

This time it's personal

The next head of the US National Human Genome Research Institute will need to be equipped to deal with the scientific, political and societal challenges presented by the burgeoning era of personal genomics.

On 28 May, Francis Collins announced that he is stepping down after 15 years as head of the US National Human Genome Research Institute (NHGRI), where he helped lead the international effort to sequence the entire human genetic code (see page 710). His departure is an opportunity to take stock of his legacy and to contemplate the future of the institute he leaves at a crossroads in its existence.

Collins took up his post in 1993, four years before the NHGRI became a separate branch of the National Institutes of Health (NIH) in Bethesda, Maryland. He proved to have a rare combination of political acumen, scientific talent and administrative skill — abilities that allowed him to steer the institute through numerous challenges and reinventions while remaining about as well-liked as it's possible for an institute director to be.

In the 1990s, for example, successful completion of the public arm of the Human Genome Project was in doubt until Collins rallied its members to respond to a challenge led by Craig Venter, a former NIH scientist who was mounting a sequencing effort for commercial purposes. After the genome project was successfully concluded in 2003, some in Washington began asking whether the NHGRI was still needed. Collins made the case — rightly — that the post-genomic era would be even more exciting and challenging than the race to sequence the genome. He and the NHGRI helped prove that by launching a variety of follow-on efforts, including the International HapMap Project, ENCODE (the Encyclopedia of DNA Elements) and, most recently, the 1,000 Genomes Project — an attempt to create an atlas of human genetic variation in unprecedented detail.

Those initiatives have been possible because of technological improvements that dramatically cut the cost of sequencing — improvements for which Collins deserves considerable credit, as he made them funding priorities. The lower costs have accelerated efforts worldwide to clarify the genetic basis of evolution by filling

out the 'tree of life'; more than 100 non-human species have been or are now being sequenced.

Collins also deserves credit for making the ethical, legal and social issues of genomics a high priority at the NHGRI. He says he is particularly satisfied with recent passage through Congress of the Genetic Information Nondiscrimination Act, for which he had long been a passionate advocate. Certainly the passage was timely: cheaper sequencing has ushered in a new era of 'personal genomics', in which companies are offering all manner of tests that claim to reveal a person's susceptibility to conditions such as Alzheimer's disease.

Although Collins says he has no concrete plans for after he steps down on 1 August, speculation is rife that he could serve as NIH director or White House science adviser in the next presidential administration. In the meantime, the future of the NHGRI is more cloudy than his own. The funding situation at the NIH has been gloomy for years, with flat budgets stifling many potentially worthy projects. And with Collins gone, the NHGRI may become more of a target for politicians who feel it has run its course.

Ideally, then, Collins's successor will display both his scientific and political skills. He or she will have to continue to pursue ambitious new projects on the scale of 1,000 Genomes. At the same time, the new director will have to ensure that the implications and applications of those projects are fully explained to all concerned, from the medical profession and insurance companies to politicians and the public. Genomics is now at a point where the science and technology are moving much faster than society's ability to assimilate and make sense of the information. Bridging this gap — a task at which Collins worked very hard — will be his successor's major challenge. ■

"The National Human Genome Research Institute may become more of a target for politicians who feel it has run its course."

On the campaign trail

It is less specific policies and more the approach to science that will distinguish the next US president.

Primary season in the US presidential race is finally over. As the dust settles, it remains to be seen how much the protracted battle between Senators Barack Obama and Hillary Clinton has hurt the Democratic party, but for now politicians of all stripes are regrouping and strategizing about how to win the election.

Those who believe that science and technology issues should be at the forefront of the campaign agenda have so far been disappointed.

The notion of a 'science debate' — itself debatable (see *Nature* 451, 605; 2008) — has fizzled out. Pleas for the candidates to address research issues seem largely to have fallen on deaf ears.

Yet there can be little doubt that the next US administration will be more science-friendly than the present one. Both of the expected nominees, Obama and Republican Senator John McCain, have put forth platforms that represent major breaks from the policies of President George W. Bush. No matter who is elected, the United States will almost certainly repeal its ban on federal funding for research on human embryonic stem cells using fresh cell lines. The new president will endorse mandatory reductions in greenhouse-gas emissions and work with Congress to enact meaningful climate legislation. And new leadership appointments at key agencies can

only mean that morale at places such as the Environmental Protection Agency will improve.

So far, McCain and Obama have set out relatively different platforms on science and technology issues. For his part, Obama has yet to give a substantive speech on science issues, as Clinton did on the 50th anniversary of Sputnik. But he has adopted many of the traditional Democratic platforms, such as increasing federal funding for biomedical research and improving the jobs pipeline for young scientists. He has put a strong emphasis on the importance of technology for improving the lives of everyday Americans and their access to government. And he has been remarkably successful at using the Internet to turn his supporters into active participants in the political process. An Obama administration might well mean that more technologically savvy people will be drawn into public life — and perhaps that more young people will choose science and engineering careers.

In contrast, McCain has revealed few details of his science and

technology agenda. But in some areas he has been quite outspoken. Last week, for example, he called for a new treaty to reduce the US and Russian nuclear arsenals, suggesting he would lower the number of weapons even beyond the cuts planned by the Bush administration.

And on 12 May he outlined a detailed plan for controlling greenhouse-gas emissions — the centrepiece of which is a cap-and-trade system of the type he has been advocating since at least 2003, when he co-sponsored the first meaningful bill on that subject.

In the end, the main factor is not how Obama or McCain feels about specific science-related issues.

What American voters deserve to know is how each candidate's mind works. Does he listen to a handful of ideology-driven advisers in making key decisions? Or does he look facts in the face and base his conclusions on all the available evidence? If it's the latter, then that is the candidate who is in sync with science at a level far more meaningful than any immediate argument over research budgets or competitiveness. ■

"There can be little doubt that the next US administration will be more science-friendly than the current one."

A flood of hard data

Social scientists have a new handle on group behaviour — but its causes remain a challenge.

Every human being makes choices, takes action and is affected by the environment in a way that seems utterly idiosyncratic. Yet in the aggregate, as the British philosopher John Stuart Mill put it a century and a half ago, human events "most capricious and uncertain" can take on "a degree of regularity approaching to mathematical". A case in point is the analysis of mobile-phone data discussed by González *et al.* on page 779 of this issue. It reveals just such a mathematical regularity in the seemingly unpredictable way people move around during their daily lives (see also page 714).

As remarkable as this result is — and it is still not completely understood — the research is just as notable for its methodology. Social scientists have long struggled with a paucity of hard data about human activities; people's self-reporting about their social interactions, say, or their movement patterns is labour-intensive to collect and notoriously unreliable. In this case, the researchers obtained objective data on individuals' movements from mobile-phone networks (albeit without access to any individual's identity, for privacy reasons). This gave them a data set of proportions almost unheard of for such a complex aspect of behaviour: more than 16 million 'hops' for 100,000 people. The resulting statistics show a strikingly small scatter, giving grounds for confidence in the mathematical laws they disclose.

The mobile-phone technique is simply the latest example of how modern information technologies are giving social scientists the power to make measurements that are often as precise as those in the 'hard' sciences. By analysing e-mail transmissions, for example, or doing automated searches of publication databases,

social scientists can collect detailed information on the network structure of scientific collaborations and other social interactions. And by allowing their subjects to interact online, researchers can do large-scale studies of, for example, the role of social interactions in opinion formation, complete with control groups and tuneable parameters.

It's not an overstatement to say that these tools are fostering a whole new type of social science — with applications that go well beyond the conventional boundaries of the field. There is sure to be commercial interest in the detailed patterns of usage for portable electronics, for example, and the nature of mass human movement could inform urban planning and the development of transportation networks. Epidemiologists, meanwhile, will no longer be forced to work with highly oversimplified models of infection rates and disease spread: recent work has clarified how the transmission of disease depends on the precise structural details of the network of person-to-person contacts.

For all their promise, making sense of these new data sets requires a rather different set of statistical skills than those needed in conventional social science — which may be one reason why studies such as that by González *et al.* are so often conducted by researchers trained in the physical sciences. To some extent this 'physicalization' of the social sciences is healthy for the field; it has already brought in many new ideas and perspectives. But it also needs to be regarded with some caution.

As many social scientists have pointed out, the goal of their discipline is not simply to understand how people behave in large groups, but to understand what motivates individuals to behave the way they do. The field cannot lose focus on that — even as it moves to exploit the power of these new technological tools, and the mathematical regularities they reveal. Comprehending capricious and uncertain human events at every level remains one of the most challenging questions in science. ■

"The goal of social science is not simply to understand how people behave in large groups, but to understand what motivates individuals to behave the way they do."

RESEARCH HIGHLIGHTS

BIOTECHNOLOGY

Genetic dexterity

Nature Biotechnol. doi:10.1038/nbt1409; doi:10.1038/nbt1398 (2008)

Two groups have developed a method to pick and choose the genes they disrupt in zebrafish using chimaeric enzymes called zinc-finger nucleases.

These enzymes cause double-stranded breaks in DNA when two of them join — and because they contain several zinc fingers that are each specific to a different triplet of DNA bases, enzymes can be engineered to target unique sequences in a genome.

Sharon Amacher at the University of California, Berkeley, and her colleagues used nucleases with four zinc fingers and observed no DNA breaks other than those they intended. The other team, led by Scot Wolfe and Nathan Lawson at the University of Massachusetts Medical School in Worcester, used three-finger nucleases and found off-target cleavage in 1–5% of zebrafish.

PLANT SCIENCE

Mossy management

Science **320**, 1181 (2008)

Careful measurement of the fixed nitrogen in water that drips through the foliage of boreal forests (pictured below) at a dozen sites in northern Sweden has helped researchers pin down how that biome manages its nitrogen budget.

Shortly after forest fires, canopy dripwater delivers sufficient nitrogen to provide for cyanobacteria on the forest floor. The cyanobacteria thus reduce the rate at which they fix nitrogen from the air, Thomas DeLuca of the Wilderness Society in Bozeman, Montana, and his colleagues have found.

Because the 12 sites had undergone varying periods of post-fire recovery, the authors were able to show that cyanobacteria increase their nitrogen fixation rates as the feather mosses they live on grow back, and as the nitrogen

Blow by blow

Nature Geosci. doi:10.1038/ngeo208 (2008)

Wind measurements have verified that the tropical upper troposphere is warming, and doing so most strongly just below where it meets the stratosphere.

The absence of this warming on decadal timescales as recorded by thermometers on weather balloons and by satellites has been a mystery to climate modellers and seized upon by global-warming sceptics. Robert Allen and Steven Sherwood of Yale University in New Haven, Connecticut, calculated horizontal gradients in temperature from vertical wind gradients since 1970. The temperature record is more muddled by non-climatic artefacts than that of wind because the



BLICKWINKEL/ALAMY

temperature-measuring instruments used are difficult to calibrate and have varied in quality over the years. The

pattern the authors found agrees with expectations derived from current climate models.

content of rainwater dips. This feedback mechanism may provide some buffer to nitrogen pollution from human sources.

EVOLUTION

Genetic bric-a-brac

Science **320**, 1210–1213 (2008)

Many single-celled organisms collect genes from other organisms — a process known as horizontal gene transfer — but multicellular organisms tend not to. Now tiny invertebrates called bdelloid rotifers have been found to take on genetic material from a range of other species, including bacteria, fungi and plants.

Multicellular creatures rarely do this because their germ line is sequestered in the gonads, explain Eugene Gladyshev, Matthew Meselson and Irina Arkhipova at Harvard University in Massachusetts. Bdelloid rotifers are different. They often experience desiccation, potentially opening up their cell membranes to chunks of outsider DNA. This unusual way of injecting diversity into their genomes may help to explain why these rotifers have gone 40 million years without sex.

ASTRONOMY

Meal deal

Astronom. J. **135**, 2287–2290 (2008)

Astronomers should, in principle, be able to spot a planet being gobbled up by its star, according to the calculations of Alessandro Massarotti of Stonehill College in Easton, Massachusetts.

In the later phase of life, most stars swell to a size large enough to swallow an orbiting planet. In some cases, the meal should be detectable as a change in the star's rotation, but astronomers have had no idea how often the circumstances necessary for observing this will arise.

Massarotti used a public database of extrasolar planets to calculate that watching planet gobbling should be possible in about 1% of cases. He believes that this result backs up three possible detections made by him and others during a survey published in 2007.

CHEMICAL BIOLOGY

Reducing power

Nature Chem. Biol. doi:10.1038/nchembio.92 (2008)

Mammals can reduce nitrate ions (NO_3^-) to nitrite ions (NO_2^-), a trick that only bacteria were thought able to perform.



J. BRANDENBURG/MINDEN PICTURES/FLPA

Jon Lundberg of the Karolinska Institute in Stockholm and his colleagues had set out to study the importance of natural gut bacteria in nitrate metabolism using mice bred to harbour no microorganisms. But when these germ-free mice were fed sodium nitrate, and nitrite ions then showed up in their blood, the team began hunting for enzymes to explain the result.

They recorded nitrate-reducing activity from xanthine oxidoreductase in the liver tissue of both rodents and humans, and found that NO_2^- can be further reduced to NO . The pathway ramped up during an experiment in which the researchers clamped the abdominal aorta of rats — perhaps unsurprisingly, given that NO dilates blood vessels.

NANOTECHNOLOGY

Sheet change

Nanoletters doi:10.1021/nl0808132 (2008)

Researchers have worked out how to detect trace amounts of amyloid- β protein that has undergone the conformational change seen in people with Alzheimer's disease.

Gerard Côté of Texas A&M University in College Station and his co-workers built a nanofluidic device able to concentrate amyloid- β proteins and gold nanoparticles at the entrance of a tiny channel by exploiting capillary flow. They then performed surface-enhanced Raman spectroscopy on the concentrated amyloid- β proteins, searching out any with a structure predominantly composed of β -sheets, a common folding arrangement.

This may allow doctors to test the cerebrospinal fluid of patients with cognitive decline and identify those who will go on to develop Alzheimer's disease, the authors say.

CHEMICAL SENSING

Molecular mapping

Angew. Chem. Int. Edn doi:10.1002/anie.200801516 (2008)

A family of molecular proton sensors that can be programmed to sit at specific distances from the surface of a membrane has been devised by Seiichi Uchiyama and Prasanna de Silva at Queen's University in Belfast, UK, and Kaoru Iwai at Nara Women's University in Japan.

These molecules have position-tuning groups that 'seek out' a local environment that matches their own compatibility with water. They thus distribute themselves at various distances along a radial coordinate of a membranous sphere called a micelle. The local proton concentration determines

the intensity of emission from a fluorescent 'reporter' group on the sensor, whereas the asymmetry of local electronic-charge polarity determines the emission wavelength.

All of these details can be mapped. This technique might one day provide clues about how biological surfaces and structures function.

ZOOLOGY

Mid-ocean wanderer

Biol. Lett. doi:10.1098/rsbl.2008.0147 (2008)

Basking sharks (*Cetorhinus maximus*; pictured below) thought to exist in discrete populations thousands of kilometres apart may in fact mix by migration, say scientists who tracked a female across the Atlantic.

Previous tagging experiments have shown



apparently distinct populations of basking sharks migrating south for winter at shallow depths along the continental shelf of Europe and the east coast of North America.

But a basking shark tagged by Mauvis Gore of Marine Conservation International in Newton, UK, and her colleagues travelled 9,589 kilometres from the Isle of Man, UK, to a region east of the Newfoundland shelf. The shark's nights were generally spent at depths of 200–300 metres and her days at 400–800 metres, once reaching 1,264 metres.

Correction

The Research Highlight 'Growing up bigger' (*Nature* **453**, 567; 2008) said that deuterium atoms in deuterated benzene can occupy more space than hydrogen atoms in benzene owing to vibrations of the carbon-deuterium and carbon-hydrogen bonds. In fact, the effect is caused by vibrations of the entire molecules about their equilibrium positions.

D. BURTON/NATUREPL.COM

JOURNAL CLUB

Uri Alon

Weizmann Institute of Science, Rehovot, Israel

A biophysicist learns the art of hugging.

If there are two things I love, they are warm hugs and simple answers to long-standing questions. Why must proteins bend in order to bind to their partners? This bending, known as induced fit, is puzzling, because it costs elastic energy and makes the binding less tight. But all sorts of proteins show induced fit — such as antibodies that recognize viruses and regulatory proteins that embrace DNA — despite the fact that such processes would be more efficient if the protein and target fitted together like pieces of a jigsaw puzzle.

Enter two Israeli physicists, Tsvi Tlusty and Yonatan Savir. They used statistical mechanics to show that bending is a good idea if the goal is not to bind tightly but to avoid binding the wrong partner (Y. Savir and T. Tlusty *PLoS ONE* **2**, e468; 2007).

Suppose that a protein needs to bind its target molecule 'A', and to avoid binding molecule 'B', which is a bit smaller than A but otherwise similar in shape. The protein would do well to make its binding pocket a little larger than A; it would then have to bend a little to embrace A, at a small energy cost, but bend a lot to bind B. Crucially, the elastic energy required for a protein to flex rises ever more steeply with the extent of bending, just as that needed for a spring to bend rises with the square of bending. So the energy difference between binding A and binding B is greater with induced fit than it is when A is a perfect fit.

Like all fruitful theories, this one makes testable predictions. It allows researchers to hypothesize what sort of imperfect fit might best serve a particular protein so that binding to non-target molecules is minimized. An antibody that attaches to virus proteins should be able to avoid similarly shaped human proteins, for instance. So perhaps, as with people, if you really want to know whether proteins are made for each other, it's in the hug.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Russian science academy rejects Putin ally

Mikhail Kovalchuk's rise to the top position in Russian science seemed a done deal. But the general assembly of the Russian Academy of Sciences has thwarted plans for the head of its newly established division for nanosciences to become the influential academy's new president.

Russia's then president Vladimir Putin last year chose Kovalchuk to head Russia's US\$7-billion push into nanotechnology. A few weeks later, the academy's leadership appointed Kovalchuk as 'acting' vice-president for nanotechnology.

Kovalchuk is the director of the Kurchatov Institute in Moscow, Russia's premier centre for nuclear science. But as he was not a full member of the academy (only a 'corresponding' member), he could not be elected president. In preparation for the election, last week's general assembly was expected to grant him full membership. But on 28 May he failed to win the two-thirds majority vote necessary for membership, by 44 votes.

According to the academy's rules, full membership is restricted to individuals who have contributed in a ground-breaking way to the advancement of science. Kovalchuk — whose older brother, Yuri Kovalchuk, is a

successful banker with close personal ties to Putin — has earned his merits as a skilled science administrator with influential contacts in political and business circles. He has also done research, mostly in crystallography, but his achievements and publication record are not considered outstanding.

Nonetheless, observers expected that the academy's long-time president, mathematician Yuri Osipov, who was on 30 May re-elected for the fourth time, would just be a placeholder for Kovalchuk (see 'Presidential election disappoints reformists'). Osipov's rival candidate, high-

energy physicist and former science minister Vladimir Fortov, who had promised he would modernize the ageing academy, failed to win a majority in the presidential elections.

According to an unwritten but, so far, strictly enforced rule, academy leaders should not be older than 70 years. Osipov, who turns 72 in July, had last year promised to step down, but changed his mind after jurists assured him that the age provision would not be an obstacle.

Kovalchuk had already pawed the ground. He told *Nature* last year that Russia needs to rethink its approach to science, which critics

"It seemed obvious that Osipov would merely keep the position for Kovalchuk."



say suffers from prevailing Soviet structures and widespread lack of competition (see *Nature* **449**, 524–527; 2007).

"Nobody really believed that Osipov intended to remain president for very long," says biologist Mikhail Gelfand, a Howard Hughes International Research Scholar and vice-president for science at the academy's

Presidential election disappoints reformists

The re-election of Yuri Osipov as president of the Russian Academy of Sciences is a bitter disappointment for the many Russian scientists who had hoped for a regime change.

With 651 votes for Osipov in Friday's election, the 71-year-old mathematician succeeded over two younger rivals, physicist Vladimir Fortov and ecologist Valery Cheresnev, who ended up with 486 and 88 votes, respectively.

Osipov has been criticized in the past for standing in the way of reform, which many scientists think is urgently needed to restore the strength and reputation of Russia's best-known scientific brand.

In an open letter to the general

assembly, Alexander Spirin, one of Russia's most highly recognized biochemists, warned that Osipov's re-election would cement the academy's reputation for being stubbornly opposed to introducing changes such as international peer review and open competition for funding.

Fortov, who served as Russian science minister from 1996 to 1998 and who then became the first president of the Russian Foundation for Basic Research, said he would increase the fraction of research money distributed by open competition. He also planned to create closer ties with the scientific 'diaspora', the large number of scientists who emigrated from the country when

they became disillusioned about the state of Russian science. And he intended to promote the recruitment of talented young scientists to Moscow, by helping them buy apartments in Russia's incredibly expensive capital.

Osipov has not yet outlined a detailed work programme for his fourth term in office. He is said to be open to the idea of building houses for young scientists on land that belongs to the academy. Other than that, he received crucial support from Prime Minister Vladimir Putin, who told the assembly on 29 May that the government will sharply increase academy members' salaries. Full members can now expect their monthly bonus payments to

rise from around 24,000 rubles (US\$1,000) to 50,000 rubles. Rank-and-file scientists, who often get less than \$300 per month, may well come away empty handed.

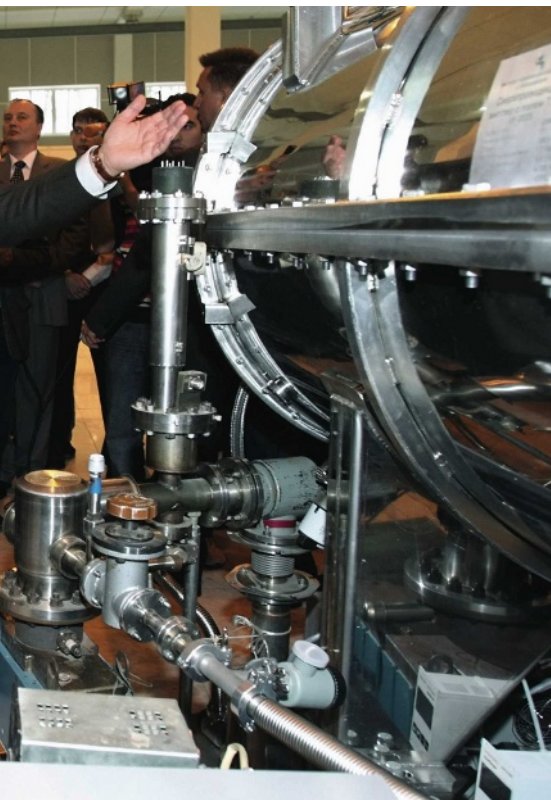
"Few 'ordinary' scientists employed by the academy, me included, would have voted for Osipov," says Konstantin Severinov, a molecular biologist at the academy's Institute of Molecular Genetics in Moscow.

Critics say that the academy is increasingly giving up its scientific ideals for power and money-making purposes. "With its many vested interests the current leadership is losing its moral standing," says Gelfand. "Fortov would have meant a real change. Alas, it was not to be."

Q.S.



PROGRAMS THAT KNOW WHAT YOU'RE THINKING
Researchers can predict which noun a person is visualizing.
www.nature.com/news



S. CHIRKOV/REUTERS

Mikhail Kovalchuk (right) has close ties with Russian prime minister Vladimir Putin (left).

Institute for Information Transmission Problems in Moscow. "It seemed obvious that he would merely act as a kind of interim president for a year or so, keeping the position for Kovalchuk."

Scientists close to the academy speculate that many members voted against Kovalchuk because they felt under too much pressure from its leadership to agree with its favourites. Others think that academics were disappointed that the money flow from Rosnanotech, a tax-exempt body set up last year to coordinate Russia's nanotechnology initiative to academy institutes, was sparser than they had hoped. Kovalchuk has a great deal of influence over Rosnanotech and is a member of its board.

Indeed, the desire for Rosnanotech money has taken strange forms. In May, an over-eager Siberian scientist sent a self-invented 'nano-powder' to at least five high-ranking academy members, including Kovalchuk, whose deputy director Svetlana Zheludeva opened the package. When Zheludeva died soon after, murder was suggested by the Russian media. Russian health officials have dismissed such speculation, saying that her death on 17 May from acute liver failure was not due to poison. ■

Quirin Schiermeier

Mars reveals early mysteries

When NASA's Phoenix spacecraft settled in Mars' northern latitudes on 25 May, there was little doubt that it would land amid a quilted pattern of troughs created over time by ice contracting and cracking below the surface. The surprise is that Phoenix is surrounded by 'polygons' much smaller than expected.

"I'm not in a panic," says mission scientist Mike Mellon of the University of Colorado at Boulder. "But I'm intrigued that there's something we can learn here about Mars that we didn't know before."

Other surprises may arise this week. As *Nature* went to press, Phoenix had scooped its first handful of martian soil and was expected to unleash a battery of geochemical instruments on soil and ice samples. But the polygons — with their direct analogies in permafrost regions on Earth — offer one of the mission's earliest mysteries.

Polygons form as winter temperatures cause the ice to contract. Eventually, the tension caused by the contraction exceeds the strength of the ice, and millimetre-scale cracks open up. Soil plugs the cracks and, when temperatures rise again, the ice buckles under the pressure, pushing up the centre of the polygons and over time creating troughs at the boundaries (see picture).

Using images from the Mars Reconnaissance Orbiter (MRO) satellite, Mellon had previously measured 915 polygons in Phoenix's landing area with an average diameter of 4.6 metres. The smallest were 2.5 metres across, although the limits of the MRO camera prevented Mellon from identifying many that were

smaller than that. When he modelled the ice-cracking process that forms the polygons, he calculated that 5 metres would be their typical diameter, he and colleagues report in a paper accepted for publication in the *Journal of Geophysical Research*. Also evident was the fainter imprint of a larger polygon system, which had troughs 22 metres apart and which could represent an earlier climatic epoch when Mars' spin axis tilted more, resulting in deeper ice near the poles.

The modelling work made perfect sense until the first images from Phoenix last week allowed scientists to measure two polygons just 1.4 metres and 2.4 metres across — significantly smaller than the 5-metre and 22-metre ones. "We have to refine our theory in such a way that it explains all three scales," Mellon says. He hopes studies of the soil and ice parameters — their depth, strength and composition — will help him to adjust his models.

On Earth, there are two main types of cracks that lead to polygons. In the wet of the Arctic, melt water from the icy edges of the cracks flows in and refreezes, building up an ice wedge over time. But in the dry deserts of Antarctica, dust falls into the cracks in the subsurface ice and builds up sand wedges.

Phoenix should be able to discern between the two crack styles by digging a trough, says Ray Arvidson, a planetary scientist at Washington University in St. Louis, Missouri, and lead scientist for Phoenix's robotic arm. Although many scientists expect to find a sand wedge, an ice wedge would be a startling discovery — evidence that water flowed near the surface in the recent past.

Arvidson served on the science team for the Viking landers in the 1970s, and one of his disappointments is that a polygon lay just out of reach of Viking 2, which, until the arrival of Phoenix, had explored the most northern latitude of Mars. This time, one polygon, dubbed Humpty Dumpty, sits directly within Phoenix's digging area.

The arm, working like a backhoe, is expected to scrape a trench from the elevated centre of a polygonal mound into a trough called the Wall. ■

Eric Hand

See Snapshot Feature, page 712.

Check our blog for updates from the Phoenix mission: <http://tinyurl.com/3hhax9>.



NASA/JPL-CALTECH/UNIV. ARIZONA

Troughs on Mars are presenting a puzzle.

Japanese nuclear plant in quake risk

A nuclear reprocessing plant in northern Japan is sited directly above an active geological fault line that could produce a magnitude 8 earthquake, some earth scientists say.

The massive Rokkasho plant for uranium enrichment, spent fuel reprocessing and nuclear-waste storage is built on an uplifted marine terrace of sloping sedimentary rock layers on the northeast coast of the island of Honshu. According to Mitsuhiro Watanabe, an earth scientist at Toyo University in Tokyo, there is an active fault lying directly under the plant. Watanabe presented his findings on 27 May at the annual meeting of the Japan Geoscience Union in Chiba.

But Japan Nuclear Fuel Limited (JNFL), which runs the plant and is based in Aomori City, disagrees, saying that Watanabe's announcement has "unnecessarily sparked fear in people". JNFL says that seismic reflection profiling shows that no part of the fault line described by Watanabe has seen any action for 1 million years, and that the fault doesn't extend beneath the plant. National guidelines issued in 2006 state that only faults with movement within the previous 120,000 to 130,000 years need be considered active when evaluating earthquake resistance of nuclear facilities. The JNFL survey concluded that there was no reason to fear an earthquake of

more than magnitude 6.5 at the site, and that the plant could withstand a 6.9 quake nearby.

Last July, Tokyo Electric Power Company's Kashiwazaki-Kariwa nuclear power plant — which was designed to withstand a magnitude 6.5 earthquake — was unexpectedly rocked by one of 6.8 (see *Nature* 448, 392–393; 2007).

Watanabe analysed JNFL's seismic reflection profiles of the Rokkasho site in addition to his own earth-deformation surveys based on aerial shots taken between February and early May this year. He says that the uplifted structure created some 120,000 years ago shows many signs of deformation since then — characteristic

of land sitting over what is called a reverse fault, which he estimates at about 15 kilometres long. "There is definitely a fault there that has been active until recently," Watanabe says. He adds that the fault might link up with an undersea fracture to create a 100-kilometre-long fault capable of pounding the Rokkasho plant with a magnitude 8 earthquake.

Seismology and earthquake-safety specialist Katsuhiko Ishibashi, emeritus professor at Kobe University, agrees with Watanabe that there is probably a 15-kilometre fault directly below the plant. The idea of a longer fault needs further investigation, he says. Either way,

Ishibashi worries that an earthquake larger than expected could inflict serious damage on the plant. "In the worst-case scenario, the whole of northern Japan and even as far as the wider Tokyo area could suffer a serious radiation disaster," he says.

Jim Mori of Kyoto University's Disaster Prevention Research Institute says that Watanabe's conclusions are reasonable, but the data could be interpreted in other ways. He recommends further study, including higher-resolution seismic surveys, bore holes drilled into the fault — which would be possible, but probably too costly — and more geological work at sites along its length.

JNFL submitted its seismic report in November 2007 to the Nuclear and Industrial Safety Agency, which is now reviewing it. An official there told *Nature* that Watanabe's critique would be taken into account, but he did not say what measures would be taken if the possibility of a larger earthquake was borne out.

The Rokkasho plant is at the heart of Japan's plan to reprocess spent fuel for plutonium that can be mixed with fresh uranium. This has met with resistance and the country has yet to decide on a site where it could build a power plant to burn the reprocessed mixed oxide fuel. The current debate is likely to complicate issues.

David Cyranoski

"There is definitely a fault there that has been active until recently."

UN decision puts brakes on ocean fertilization

Ocean-fertilization advocates suffered another setback last week as 191 nations agreed to a moratorium on large-scale commercial schemes to mitigate climate change.

The agreement, adopted on 30 May at a meeting of the United Nations Convention on Biological Diversity in Bonn, Germany, calls for a ban on major ocean fertilization projects until scientists better understand the potential risks and benefits of manipulating the oceanic food chain. It took 12 days of diplomatic effort to win the support of Australia, Brazil and China, which had opposed the moratorium.

Fertilization projects typically involve seeding the ocean with some form of iron to stimulate algal

growth. The algae absorb carbon dioxide from the atmosphere during photosynthesis. A number of ocean-fertilization companies have been formed, anticipating that they could sell credits for carbon-dioxide reductions into carbon markets, but many scientists question the effectiveness and potential side effects of fertilization.

The new language endorses and broadens a similar warning given last November by 35 countries party to the London Convention and Protocol, which governs ocean pollution. Those nations agreed to study the issue and to establish rules this year.

Dan Whaley, who heads Climos, a San Francisco-based company promoting ocean fertilization,

declined to speculate on the impact of last week's decision. But he endorsed the call for more science and said he fully supports efforts to address the issue under the London Convention.

Advocates claim that algal blooms will briefly flourish, then die and drop to the bottom of the ocean, taking atmospheric carbon with them. But there are questions about how much of the algae might be consumed by other organisms or be broken down before sinking, reducing the amount of carbon sequestered. And altering the ocean's ecology may have unwanted effects, such as increased acidity or decreased oxygen levels.

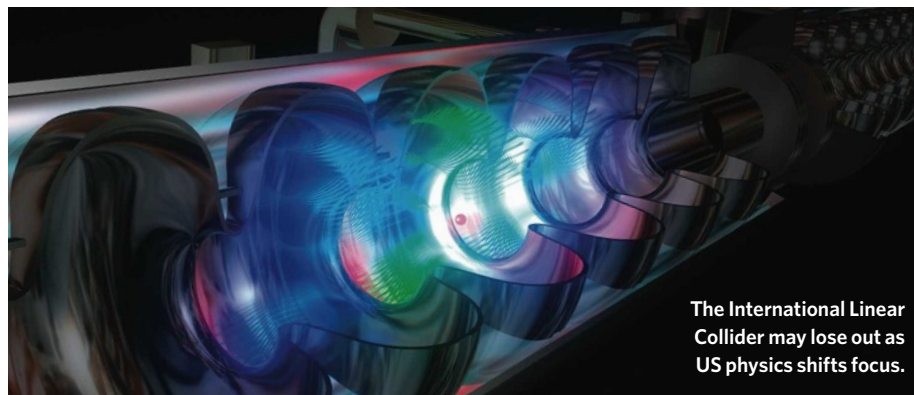
Biological oceanographer Mike Behrenfeld at Oregon State University in Corvallis says the

moratorium is justified. "We have no idea how long-term, sustained iron fertilization is going to influence the species composition."

But the only way to find out is through the large-scale experiments that commercial interests are best poised to carry out, says Ken Johnson, a senior scientist at the Monterey Bay Aquarium Research Institute in California.

Johnson is against allowing companies to market carbon credits yet, but he thinks that ocean fertilization is the most viable geo-engineering option for addressing a runaway climate. "This isn't something to rush into, but it's the only solution we've got if climate gets out of control."

Jeff Tollefson



D. HAMBURG

The International Linear Collider may lose out as US physics shifts focus.

Physicists to target neutrinos

The new focus for America's high-energy physics should be an elusive one: the zippy, chargeless, near-massless neutrino, according to a report that provides the US Department of Energy (DOE) and the National Science Foundation (NSF) with a roadmap for the next decade.

The report, written by the Particle Physics Project Prioritization Panel (P5) and endorsed on 29 May by a DOE and NSF advisory committee, suggests that the US physics programme should concentrate on the fertile terrain of neutrino physics rather than focusing all its efforts on the high-energy frontier explored by colliders.

The report also seems to spell the end for US ambitions to host the International Linear Collider (ILC), a planned smasher of electrons and positrons that would complement the Large Hadron Collider (LHC), which is to begin operations this summer at CERN, the European particle-physics laboratory near Geneva. Some physicists wanted to pursue an aggressive schedule for the ILC. But until the LHC starts producing results, it will be unclear whether the ILC has been designed to be powerful enough.

Yale University's Charles Baltay, who chaired the P5 panel, says that research and development should continue for the ILC, but that it was no longer "front and centre". "This was obviously the thorniest issue the panel faced," says Baltay. "We spent four months arguing about this."

Instead, the P5 report recommends more work on the neutrino, which was discovered to have mass only a decade ago. Physicists want to understand how the three known neutrino 'flavours' morph from one type to another. Some physicists hope that by understanding neutrinos and their

antiparticles, they will be able to explain why the Universe ended up being made of matter rather than antimatter.

To this end, the P5 report recommends starting construction of NOvA, a programme that would send a neutrino beam from Fermilab in Batavia, Illinois, to detectors in an underground mine in Minnesota. The report endorses also building a more powerful neutrino beam that would travel farther, from Fermilab to the planned Deep Underground Science and Engineering Lab (DUSEL) in the Homestake Mine of South Dakota. The farther the neutrinos travel, the more sensitive the experiments are.

But the future of the ILC, at least for it being hosted by the United States, looks dim. The recommendations to continue ILC research and development are "okay", says Barry Barish, a high-energy physicist at the California Institute of Technology in Pasadena, who heads the ILC design effort. "But it isn't an aggressive programme to bring the ILC to the United States." Just last year Barish spoke of starting

ILC construction by 2012; now he just wants to have a construction proposal by 2012. Japan is now the front runner for hosting the project, Barish says, with CERN also being a contender after it finishes planned upgrades to the LHC.

Given the sharp budget cuts for high-energy physics in 2008, there is no guarantee that the US Congress will pay any attention to the physicists' roadmap, which considers four budget scenarios. "What happens inside Washington happens," says Melvyn Shochet, a physicist at the University of Chicago, Illinois, and chair of the DOE and NSF advisory committee. "But you need to be able to make the case."

Eric Hand

"We spent four months arguing about this."

ON THE RECORD

"Clearly, having a working toilet is a priority for us."

Shuttle payload manager Scott Higginbotham, on the failure of the toilet system on the International Space Station. The three astronauts aboard used the one on the moored Soyuz capsule instead.

SCORECARD



Skydiving balloon

A \$200,000 helium balloon, supposed to take French daredevil Michel Fournier 40 kilometres up for a record-setting sky-dive, launched successfully last week — but without Fournier. A technical problem led to its premature release. It was Fournier's third failed attempt at the jump, but he says that he'll try again in August.



Party balloons

The California Senate has voted to ban helium-filled metallic balloons because they, too, frequently escape, and can get tangled in electrical lines. They have apparently caused hundreds of power outages in recent years.



NUMBER CRUNCH

830 years ago the Tower of Pisa first began to lean. It has now officially stopped moving.

28 years ago, the latest attempts to un-tilt the tower began.

70 tonnes of earth were shifted to prop up it up.

300 years is how long it is expected to remain stable at its new lean, which is about the same tilt it had in 1700.

Sources: NASA, Globe and Mail, Mercury News, Reuters

D. STOBBE, THE CANADIAN PRESS/AP

K. SMITH/ISTOCK/ISTOCKIMAGES

The four-year fight for biological art

Steven Kurtz, an art professor at the State University of New York at Buffalo, was cleared in April of wire and mail fraud charges, four years after the FBI seized art supplies from his home that included laboratory equipment and bacterial cultures. Kurtz is a founding member of the Critical Art Ensemble, which recently exhibited an installation called *Immolation* at the Foundation for Art and Creative Technology in Liverpool, UK, showing the effects of incendiary weapons on samples of human skin.

How does it feel now the court case is finally over?

It's really kind of odd. You would think that I would be feeling light as a feather and dancing down the street. But quite the opposite, it's more like having some kind of post-traumatic stress disorder. It's only in my more practical mind that I'm very happy about it all. After four years it's hard to come down out of fight mode, and it's hard to believe that it's over.

What were you working on when you were arrested?

We had just completed a piece called *Free Range Grain* where we tested unlabelled food in Europe to see if it was free of GMOs [genetically modified organisms]. We were coming back to do it in the United States, where we were going to test labelled organic foods. We were hoping to elicit further conversation about what it would mean if, say, a handful of corporations controlled the food supply. We were ready to open the project at Mass MoCA [Massachusetts Museum of Contemporary Art in North Adams] when I was arrested and all my materials were confiscated by the FBI.

The second thing that we were doing was investigating the history of germ warfare and why in most countries the plug has been pulled. We were just interested in presenting to the public a more rational history of it, rather than this kind of hysterical response that the Bush administration threw on the public.

You can see why the FBI wouldn't be pleased with either of these things.

The second project you mentioned, *Marching Plague*, involved throwing benign bacteria at guinea pigs, referencing an old British military experiment.

Yes, it did.

Has the US Department of Justice returned any of your materials?

That's the next fight for me: to try to get my lab equipment back and other things that they took, like my computers and books.



The government alleged that you and your colleague Robert Ferrell exchanged bacterial cultures in an unauthorized manner. What happened?

[Robert] did provide us with bacterial samples. This is the most ironic thing of the whole case, because we could have bought them ourselves. When we use bacteria in our projects, it often makes cultural institutions nervous because they don't know about bacteria. We knew this, and we thought that the best way to calm the nerves of cultural institutions that don't generally have a scientist consultant on board was if we could provide them with one. We were trying to make everyone feel safer and even get rid of the perception that there could possibly be a danger, which is what ended up actually getting us in trouble.

Has the high-profile nature of the case affected your collaboration with scientists?

For our more recent projects we've either gone to Europe to do them or to Australia. In a way, our practice got expatriated a little bit after this mishap. The FBI was very thorough about going round to all the cultural institutions and labs we worked with and intimidating and threatening them. It had the effect of almost classifying science to

make sure it's further alienated and pushed away from the public.

It sounds like your art is about science education.

One of the things we're really interested in is being some kind of mediator — some type of conduit — between the sciences and the public who are outside the sciences. If we're going to understand how the results of science are going to be placed into political economy and want to have some accountability there, the only way that can be done is through informed citizenship. And we want to promote that.

Will you continue to work on biologically inspired art?

We're very interested in continuing the line of anti-war work. That's probably going to be our main subject, but we always seem to somehow fold back in, or intersect back in, life sciences' role in the war economy.

How would you characterize the climate in the US art community after 9/11?

It has just been horrible. We're talking about the government pressure on doing activities that are perfectly legal. We're talking about activities that at one time in this country were considered a great idea, like amateur science. Say someone wants to have a laboratory in their house. I guess not a lot of people have that any more, so that's weird. And because it's weird, it's suspicious. And because it's suspicious, we have to think it's probably terrorism.

How do you function culturally if that is the structure? All the FBI is protecting us from in that way is stuff that's weird, not terrorists. And if you're an artist, your job is not to be a part of the everyday and the usual. You're supposed to go to the opposite end to find new ways of thinking about culture and its images, to try to reframe issues in new and exciting ways. And in so doing you're going to do weird stuff. If weird is suspicious and suspicious is going to be equated with terrorism, then the arts and culture are in a lot of trouble — and we are.

Interview by Rachel Courtland.



Humphrey Bogart
and Ingrid Bergman
in the 1942 enduring
classic *Casablanca*.

THE KOBAL COLLECTION/WARNER BROS

Here's looking at you, kid

"What in Heaven's name brought you to Casablanca?" embittered hero Rick Blaine (Humphrey Bogart) is asked early on in the movie. The same might be asked of the millions of people who continue to watch *Casablanca* more than 60 years after its first release — the movie is consistently ranked in the top ten in lists of greatest films, to the astonishment of those involved in its production.

So what made *Casablanca* such a success? The answer could be revealed by script-reading software designed to help writers create blockbusters rather than flops. Such software is in hot demand in Hollywood, and over the past few years companies have been developing tools to help producers predict box-office takings, given a script and casting and budget information.

Now, a team at Royal Holloway, University of London, UK, has developed an algorithm to keep scriptwriters ahead of the game. The algorithm assesses how good a script is as it is being bashed out on the keyboard (F. Murtagh, A. Ganz & S. McKie *Pattern Recognit.*; in the press). Computer scientists Fionn Murtagh and Stewart McKie teamed up with screenwriter Adam Ganz to find an objective way to compare new scripts with successful ones in their genre. They analysed a selection of scripts from films deemed successful by critics to find common patterns.

"On the crudest level, we could distinguish action films, such as *Die Hard*, from a drama such as *Casablanca* just on the frequency of verbs compared with character names," says Ganz. "We wanted to see if we could find a more sophisticated way to uncover deeper structural patterns."

Their algorithm analyses how often and where each word occurs. It reveals that popular screenplays, such as *Casablanca* and episodes of the television series *CSI*, share characteristic patterns, says Ganz. For instance, they build tension in waves, by shortening the length of successive scenes in blocks to create 'mini-cliff-hangers' that are then resolved in a longer denouement. The length of these blocks and the pacing of the scenes within them are specific to genre.

The algorithm also identifies how incidents that contain conflict are distributed through the plot, by analysing how the locations of groups of frequently used words change in relation to each other. Two of the most common words in *Casablanca* are the character names Rick and Ilsa, and the algorithm can track how these two words cluster together and move apart over the course of the film, says Murtagh. "It's surprising how well this maps to Ilsa's conflicting emotions in the film as she is alternately drawn to Rick and resists him."

Ganz notes that these features are well recognized by film scholars. "We haven't stumbled on a secret formula; we're showing that patterns that experts understand can easily be quantified," he says.

As they didn't have access to scripts from unsuccessful films, the team tested the algorithm by tweaking the *Casablanca* script to create versions that had the same plot but a different scene and sub-scene structure. The algorithm predicted that films tweaked in these ways would perform badly.

The team hopes to combine the algorithm with existing software that is used to

automatically format screenplays. But is there a danger that the software will kill off innovative scripts that don't fit the Hollywood mould? Ganz does not think so. Even the unconventional movie *Memento*, in which half of the plot plays out in reverse chronological order, performs well in this type of analysis, he says. "If anything, we have the potential to separate the truly innovative from films that are superficially innovative but have no deep structure."

Not everyone is convinced. "This methodology downplays the visual and sonic qualities of film-making," says Frank Krutnik, an expert in film studies at the University of Sussex, UK.

Ganz counters that although visual and sonic elements are important, producers only have a script in hand when deciding whether to green-light a project and must imagine the rest.

"Screenplays are odd things: they are assessed by being read, but if they are successful, they will never be read again," says Ganz. "Writers and script editors need all the help they can get with them."

McKie's group has produced a website at which writers can upload their scripts for preliminary analysis and look at frequently appearing word clusters in various successful movies, such as *Juno*, see: <http://scriptcloud.screenplayanalytics.com/>.

"The algorithm could be a really exciting tool for approaching the rewrite, illuminating possibilities for the tone, pace, poetics and even the structure of the narrative," says Amanda Holiday, a film and television screenwriter based in Cape Town, South Africa.

Zeeya Merali

Biological tools revamp disease classification

Just as different genes can contribute to similar diseases, so the same genes and families of genes can play a part in a range of different diseases. “The days when we had one gene, one disease have been dead for some time now,” says Nicholas Katsanis of Johns Hopkins University School of Medicine in Baltimore, Maryland.

Scientists are using structural biology, genomics and biochemistry to reinvigorate a decades-old debate about the relationship between genes and disease. “If you have the genes and the mutations, you can now see how much of your previous disease classification system was real and how much was wrong,” says clinical geneticist Han Brunner of Radboud University Nijmegen Medical Centre in the Netherlands.

“The molecular biological information gives us a much finer view of what really is clinically important.” It questions the idea that diseases with similar symptoms have the same genomic and biological origins.

A wave of studies is revealing the complicated relationship between genes and disease even in those disorders linked to glitches in a single gene. For example, a gene called *XPD*, which encodes a protein involved in DNA transcription and repair, has long intrigued scientists because different mutations in it cause three separate diseases. On 30 May, two separate teams published structures of *XPD* proteins in microorganisms^{1,2}, illuminating the causes behind *XPD*’s ‘three fates’.

The teams found that the *XPD* mutations causing one disease, xeroderma pigmentosum, damage the part of the *XPD* enzyme that binds DNA. As the mutated enzyme can’t repair the DNA damage caused by sunlight, it explains why patients with this condition have such high rates of skin cancer. But they found that mutations causing trichothiodystrophy — marked by premature ageing — block the protein from joining in a larger complex that transcribes DNA. If the complex can’t work, patients don’t make enough protein, and their cells die young. But so might any incipient cancer cells, explaining why patients with this disease don’t have the high cancer rates seen in those with xeroderma pigmentosum, the teams suggest.

But the *XPD* story has another layer of

complexity: patients with the same underlying genetic mutations sometimes show different symptoms. That situation is familiar to human geneticists such as Brunner, who says that scientists should begin paying more attention to these supposed ‘outliers’. “Just by accepting this as clinical variability, we are throwing away a lot of interesting biologically relevant information,” he says.

In March, Katsanis’s team published a paper on two diseases caused by malfunctioning cilia — tiny hairlike structures ubiquitous in the body. The paper³ showed that genes that cause one such disease, Meckel-Gruber syndrome, can also be mutated in patients with another ciliopathic disease, Bardet-Biedl syndrome. Katsanis’s team further found that

“The days when we had one gene, one disease have been dead for some time.”

when patients carry mutations in genes associated with both diseases, they have unique symptoms not seen in either condition alone.

Katsanis’s team went on to do further studies in a zebrafish model and found that, indeed, genes linked to the two conditions interact with each other during development. Katsani proposes that the two conditions actually reflect disruptions in a single biological pathway. “Once we define functional modules containing multiple genes and look at disorders whose phenotypes fit into broad dysfunction of these modules, we’re going to start seeing these phenomena time and time again,” he says. Systems biologists are already attempting to define such ‘functional modules’ by drawing maps of connections between genes and diseases that have been found in association studies.

Although the price tag for genetic studies is dropping, clinicians’ expertise remains expensive, and that worries Brunner. “Certainly with the amount of money people are spending genotyping thousands of patients, they are finding that the quality of the phenotype data is crucial,” Brunner says. That’s why, he says, a ‘phenome project’ is needed to investigate connections between phenotypes. The idea has been proposed before but stalled for lack of funding ■

Erika Check Hayden

1. Fan, L. *et al. Cell* **133**, 789–800 (2008).

2. Liu, H. *et al. Cell* **133**, 801–812 (2008).

3. Leitch, C. C. *et al. Nature Genet.* **40**, 443–448 (2008).

US genomics leader bows out from institute

M. BARTLETT/NHGRI

Francis Collins, the geneticist who led the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, through the completion of the Human Genome Project and the dawn of the personal genomics era, has announced that he will leave his post on 1 August.

In a news conference last week, Collins said he has no concrete plans about his future and was leaving the NHGRI to explore posts “that would be very difficult to consider or discuss or pursue while continuing in my role as a federal employee”. Many speculate that Collins could assume a leadership role in the next presidential administration, possibly as director of the National Institutes of Health or science adviser to the president. Collins also says that he would like to write a book about personal genomics for the public.

During his 15-year tenure, Collins led



Francis Collins.

what many scientists see as a shift towards funding large-scale projects. He became known as a public face of the Human Genome Project and drew the wrath of some scientists for arguing that science and religious faith are compatible (see *Nature* 442, 114–115; 2006).

See Editorial, page 697.

Merck scores victory in three Vioxx appeals

Appeals courts in two US states last week overturned verdicts favouring plaintiffs who had sued Merck, the pharmaceutical giant, over its painkiller Vioxx (rofecoxib). The drug was taken off the market in 2004 after a study showed it doubled the risk of heart attacks and strokes.

In Texas, a three-judge panel reversed a \$26-million jury verdict in the first and most public Vioxx lawsuit. It said that lawyers for Carol Ernst had failed to prove that the drug caused the death of her husband in 2001 (see *Nature* 436, 1070; 2005). Ernst's lead attorney said they would appeal the decision.

The ruling came the same day that a New Jersey appeals court scuttled \$9 million in punitive damages awarded to John McDarby in 2006 — arguing that the law under which the money was awarded is trumped by the

federal law under which the Food and Drug Administration approves drugs for market.

The reversals mean that, out of 16 Vioxx cases that have completed trials so far, only three have resulted in unqualified victories by plaintiffs. The bigger picture for the company and for tens of thousands of other plaintiffs remains largely unchanged because of a \$4.85-billion settlement agreed last November (see *Nature* 450, 324; 2007).

Indian coral islands under threat from algae

A dispute about non-native algae has broken out in India between beverage giant PepsiCo and the Central Salt & Marine Chemicals Research Institute (CSMCRI), which is based in Bhavnagar, Gujarat.

Institute researchers originally imported the alga *Kappaphycus alvarezii* for research; in 2001 PepsiCo began cultivating it for the food thickener carrageenan in the Gulf of Mannar marine bioserve, along India's southeastern coast. Reporting in *Current Science* last month, scientists from Thiagarajar College in Madurai say that corals fringing an island in the gulf are being smothered to death by the algae.

The CSMCRI suspects that the algae drifted into protected waters from PepsiCo's



S. CHANDRASEKARAN/THIAGARAJAR COLL. **The alga *Kappaphycus alvarezii* is choking coral.**

cultivation sites, whereas PepsiCo says the more likely source is the institute's seaweed depot on the island. According to CSMCRI, that depot was closed down in 2003.

Ecologists are worried that *K. alvarezii*, which is currently spreading asexually, could switch to sexual reproduction by spores. These could be carried by wind to the remaining 20 coral-fringed islands in the bioreserve.

Donation breathes life into Fermilab's balance sheet

A private donor has given US\$5 million to particle physics. The donation ends unpaid leaves of absence that physicists at Fermi National Accelerator Laboratory in Batavia,

Illinois, have been forced to take since February because of budget cuts.

The money, from a family that wishes to remain anonymous, was given on 27 May to the University of Chicago, which in turn will hire Fermilab to do contract work on neutrinos and rare particle decay. The furloughs, scheduled to last into September, were intended to save \$12 million. The \$5 million gift, plus an additional \$1 million of savings through early retirements, allowed the leaves to end on 31 May.

Fermilab has also reduced from 200 to 140 the number of lay-offs it expects to make starting in June.

Researchers kidnapped near Atacama telescope

Officials at the Atacama Large Millimeter Array (ALMA) telescope are stepping up security at its construction site in Chile after two researchers were kidnapped on a nearby road.

On 11 March, Rolf Güsten, an astronomer with the Atacama Pathfinder Experiment, an ALMA-related project, and a senior engineer were stopped by four men dressed as policemen on a highway near the site in a remote part of northern Chile. The men commandeered the vehicle and drove it into

Bolivia, where a few hours later they left the pair at the side of the road. "Fortunately, none of us was injured," Güsten says.

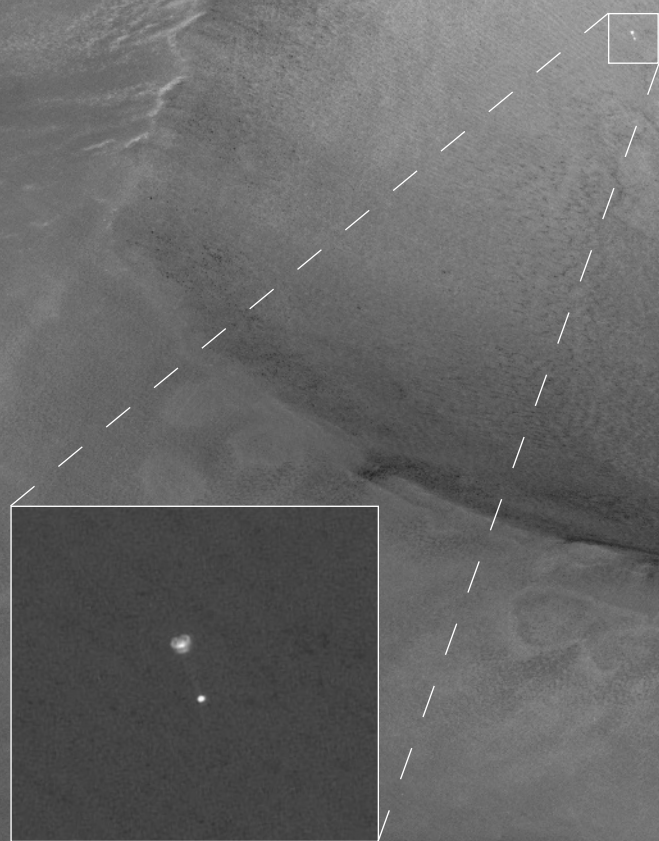
ALMA has since increased security by hiring more guards and adding more cameras and phones at the site, and is working with Chilean authorities.

Correction

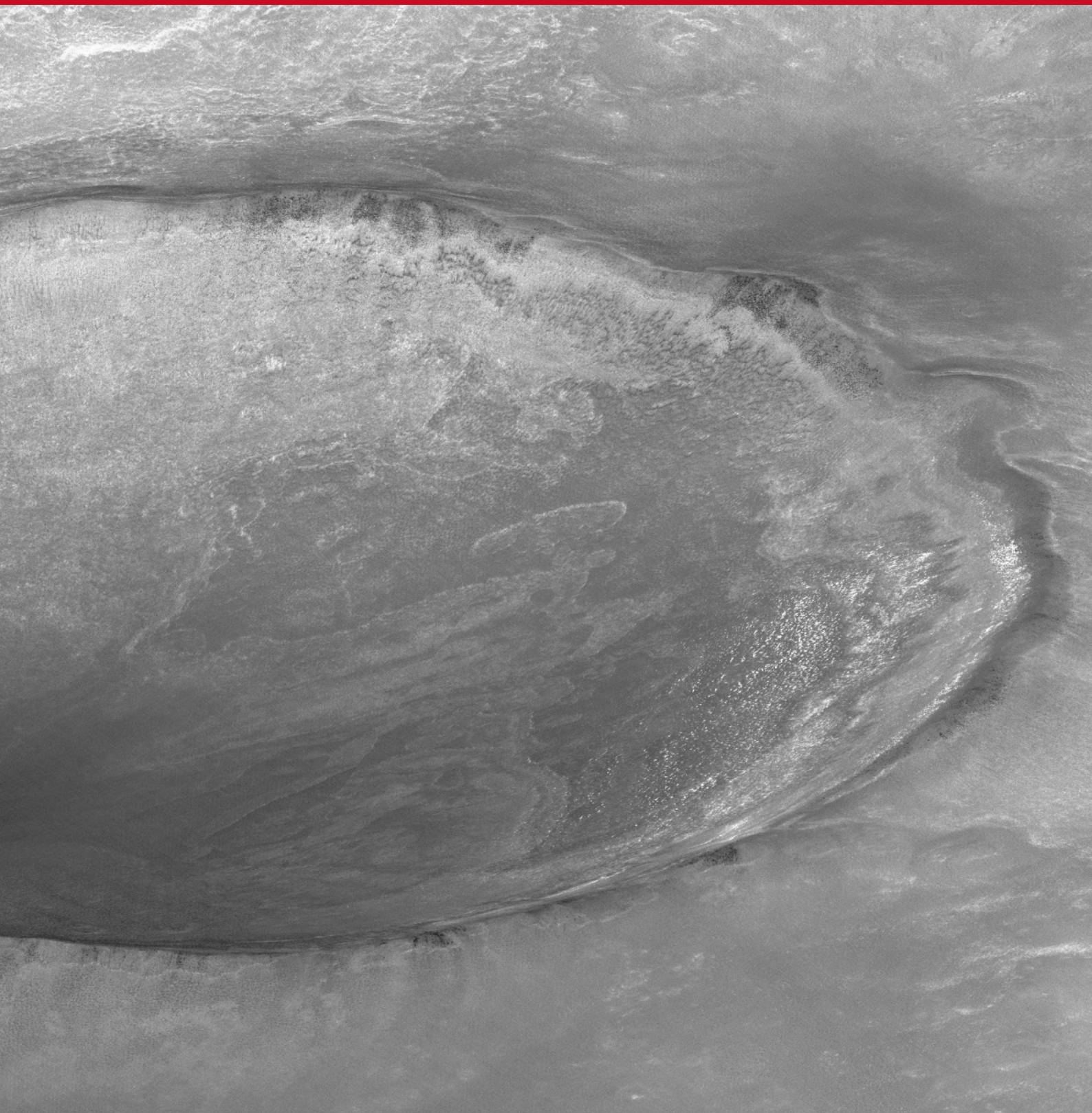
The Editorial 'Working double-blind' (*Nature* **451**, 605–606; 2008) referred to a study¹ that found more female first-author papers were published using a double-blind, rather than a single-blind, peer-review system. The data reported in ref. 1 have now been re-examined². The conclusion of ref. 1, that *Behavioral Ecology* published more papers with female first authors after switching to a double-blind peer-review system, is not in dispute. However, ref. 2 reports that other similar ecology journals that have single-blind peer-review systems also increased in female first-author papers over the same time period. After re-examining the analyses, *Nature* has concluded that ref. 1 can no longer be said to offer compelling evidence of a role for gender bias in single-blind peer review. In addition, upon closer examination of the papers listed in PubMed on gender bias and peer review, we cannot find other strong studies that support this claim. Thus, we no longer stand by the statement in the fourth paragraph of the Editorial, that double-blind peer review reduces bias against authors with female first names.

1. Budden, A. E. *et al. Trends Ecol. Evol.* **23**, 4–6 (2008).

2. Webb, T. J., O'Hara, B. & Freckleton, R. P. *Trends Ecol. Evol.* doi:10.1016/j.tree.2008.03.003 (2008).



STRANGER IN A STRANGE



LAND

On 25 May, NASA's Phoenix lander was 10 kilometres above the surface of Mars and less than three minutes from landing when it was captured passing in front of Heimdall crater by the HiRISE camera on the Mars Reconnaissance Orbiter (MRO). The two spacecraft were moving with a relative velocity of about 3.4 kilometres a second at the time, and were as far apart as Paris and Dublin.

The picture almost didn't happen. Engineers worried that HiRISE would interfere with the radio on the MRO that was tracking Phoenix. The clinching argument was that, if things went wrong, the image would show whether the lander's parachute (inset) had unfurled properly or not. By the time the image data were returned to Earth the lander was known to be safe and sound and getting ready for science (see page 703). The HiRISE image wasn't needed for a post-mortem and could be appreciated for what it is: a remarkable testament to human achievement. **Eric Hand**

NASA/JPL/UNIV. ARIZONA



Animal behaviour is an endless challenge to mathematical modellers. In the first of two features, **Mark Buchanan** looks at how a mathematical principle from physics might be able to explain patterns of movement. In the second, **Arran Frood** asks what current models can teach us about ecological networks half a billion years old.

THE MATHEMATICAL MIRROR TO ANIMAL NATURE

Food is not, in general, spread equally around the world; it comes in lumps. Foragers thus need a strategy for finding those lumps. One appealing option is a Lévy flight — a mathematical concept used in physics. Lévy flights are many-legged journeys in which most of the legs are short, but a few are much longer. They are found in some sorts of diffusion, in fluid turbulence, even in astrophysics. In animal behaviour, the longer the flight, the farther afield a creature will get, offering a way to efficiently exploit food nearby but also to discover sources farther away.

"The pattern captures what biologists often

notice," says behavioural ecologist David Sims of the Marine Biological Association Laboratory in Plymouth, UK. "Animals often take lots of short steps in a localized area before making long jumps to new areas."

But just because it makes qualitative sense doesn't mean it is a mathematical key to the real world. Hard evidence is needed to show that the pattern is a real Lévy flight, in which the frequency of steps of given distances is firmly constrained. And this evidence is what physicist Gandhimohan Viswanathan, then a graduate student at Boston University in Massachusetts, and his colleagues seemed to find in 1996.

Albatrosses soar over tremendous distances as they circle the oceans, alighting here or there to feed on squid, fish or krill before heading off again. Observers had thought the foraging was random; but any hidden pattern would be evident only on the scale of seas and oceans. It was this large pattern that Viswanathan, now at the Federal University of Alagoas in Brazil, decided to look for, using electronic logging data gathered by field ecologists at the British Antarctic Survey (BAS) in Cambridge.

Viswanathan and his colleagues found a scale-free fractal-like pattern in the data¹, just what a Lévy flight ought to produce. Three years

C. BERRIE

later, they seemed to be on the track of a new principle of ecology when they showed that this way of moving is, under some conditions, theoretically the best way for animals to find scarce prey². They and other researchers soon reported the same pattern in the movements of everything from reindeer and bumblebees to soil amoebas and the habits of fishermen³. The phenomenon is attracting more and more interest, and it seems to apply to more than just foraging. Research in this week's *Nature* shows that it applies to the movements of mobile-phone users⁴ too (see page 779).

Flights of fancy?

There's just one problem. Although other examples stand up to scrutiny, the one that started the field off does not, at least for now. There's a lesson in that. When modellers use data from the field, they have to be sure that the data really represent what they think they represent, and that they fit tightly to their model. The devil is in the detail, when sparse data can put almost all conclusions on shaky ground.

The case for Lévy flights by albatrosses ran into problems in 2004, when physicist Sergey Buldyrev, also of Boston University and one of Viswanathan's co-authors on the original albatross paper, analysed new data on albatross movements. The Lévy pattern didn't turn up. Revisiting the original data collected by the BAS researchers, Buldyrev discovered that the longest flights recorded, which were crucial to the distinctive fractal fingerprint, might have been artefacts of the recording technique.

The original albatross data came from devices called immersion loggers attached to the birds' legs. The devices recorded the proportion of time in each quarter-minute that the birds sat on the sea surface. From these data, the researchers could then infer flights as periods during which the birds remained dry. From five birds, the researchers had obtained a total of 363 flight times, which seemed to show the Lévy pattern.

But Buldyrev wondered whether the longest periods of dry-leggedness — which always seemed to be the first and last in a bout of movement — might in fact record the birds sitting on their nests. The data had not been saying what the team thought they were saying. Finding that the Lévy pattern vanished when these data points were omitted, Buldyrev and his colleagues wrote up a manuscript and sent a draft to ecologist Richard Phillips at the BAS. Phillips, working with ecological modeller Andrew Edwards, also at the BAS, confirmed that there was no support for Lévy flights. Later, when they discovered that some of the albatrosses also had location trackers fitted to them, the BAS team proved that the birds

weren't moving during the alleged long flights. "I was disappointed," says Viswanathan, "but also curious, surprised and perplexed."

The Lévy flight notion took another blow last October, when the Boston and Cambridge groups collaborated to publish a comprehensive reanalysis of the original albatross data, including an analysis of a new data set and a reconsideration of earlier studies of deer and bumblebees⁵. They found that the deer and bumblebee data were also ambiguous — the deer data, for example, actually reflected time spent cropping and processing food at a particular feeding site, rather than time spent moving between sites. Using improved statistical techniques, the teams found that none of the data offered strong support for the Lévy flight pattern. The results, they say, "question the strength of the empirical evidence for biological Lévy flights."

It looked like a simple tale of problematic data corrected. But later last year, Sims and his colleagues presented strong evidence for Lévy-like patterns in the foraging of numerous marine predators, including sharks, turtles and penguins⁶. They used what all researchers agree are more sophisticated statistical methods, and much larger data sets. Sims and others now suggest that the data really do point to Lévy flights for a variety of animals, including humans.

Not everyone yet agrees with this position. But they do agree that the episode illustrates the difficulties inherent in identifying statistical patterns with limited data. The difference between a Lévy flight and a more familiar form of random walk, brownian motion, is the distribution of steps of different lengths. In brownian

motion, as seen in the jittering of a pollen grain buffeted on all sides by invisible molecules, the distribution of distances follows a bell-shaped curve, so the size of the next step is at least crudely predictable — it is never 10 or 100 times bigger than the average, for example.

Doing the Lévy walk

A Lévy flight is a similar sort of random walk — but the distribution of distances is different. For example, the probability of large steps of size D might fall off in proportion to $d^{-\gamma}$, with γ being a number somewhere between

1 and 3. This distribution, in what is known as a power law, gives more frequent long steps than a bell curve, and produces a pattern characterized by lots of smaller movements broken episodically by long excursions.

Diagnosing a true Lévy flight means showing that the power-law distribution holds. There is a simple statistical approach to

this. First 'bin the data': that is, count up the events that fall within each small range of distances to get a measure of the way the probability of differing distances is distributed. If a power law holds, the relationship between the logarithm of this probability distribution and the logarithm of the distance will be linear. Hence, if the log of the first is charted against the log of the second, you'll get a straight line.

As Edwards points out, however, this technique can lead to trouble. "It's well known that log-log axes tend to make relationships look straight." The problem is at its worst when data are in short supply. A more rigorous approach, he says, is to decide mathematically which of two possible distributions, say a power law or an exponential, the data fit better. But such

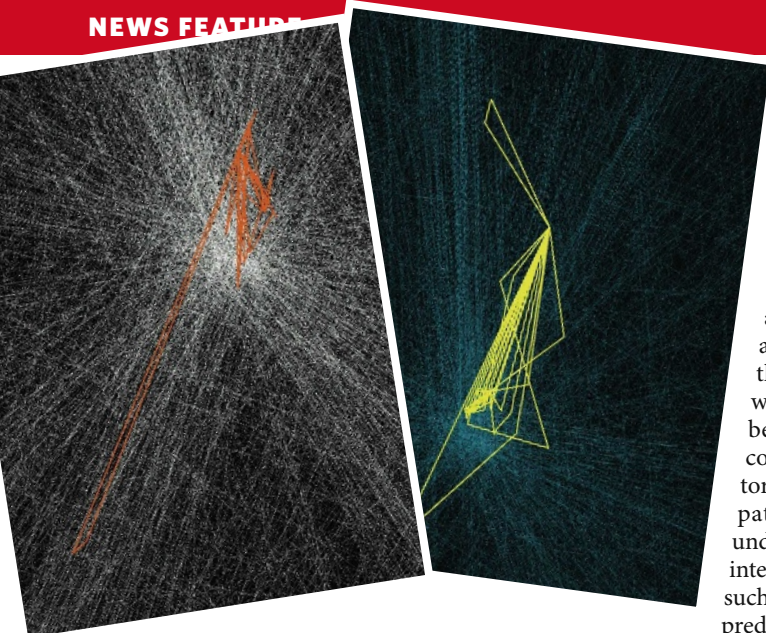
"It was wrong, yet it turned out to be fruitful because it led to other ideas that we now think are correct."

— Eugene Stanley



Dry of foot, but not flying.

T. ALLOFS/ZEFA/CORBIS



Mobile phone records reveal peoples' movements.

determinations need a lot of data.

Sims agrees. Inspired by the original albatross paper, he and his colleagues used satellite-linked tags to gather data on plankton-feeding basking sharks. They found horizontal tracks reminiscent of Lévy-like movements, but never obtained enough data to permit a sound statistical analysis. "A lack of data," he says, "means you can fail to detect the pattern even if it's there, or detect an apparently similar pattern even if it is not."

Two years ago, Sims hit on the idea to look at sharks' vertical movements instead. These were recorded at 1-minute intervals for months on end, providing more than 400,000 data points for analysis. Using statistical methods developed in part by Mark Newman of the University of Michigan, Ann Arbor, and similar to those used by Edwards and his team, they found a strong signal of Lévy behaviour. Sims then organized a collaboration of 18 researchers from four countries to gather and test similar data for other marine predators, finding the Lévy pattern for tuna, cod, leatherback turtles and penguins⁶.

Sims says that his paper "represents some of the strongest evidence for Lévy-like behaviour in wild predators." "The debate has shifted," says Frederic Bartumeus of Princeton University in New Jersey, who in 2003 found Lévy patterns in the movements of plankton. "The question now isn't whether animals perform Lévy walks, but when they do — and why?"

Although welcoming the use of larger data sets, Edwards, now at the Pacific Biological Station in Nanaimo, British Columbia, Canada, doesn't think that these studies end the debate. He says that some scientists have started to use the somewhat softer phrase 'Lévy-like' to describe their results, which may make their claims more defensible, but also introduces some vagueness into the discussion. "How 'non-Lévy-like' do the data have to be for them not to be considered 'Lévy-like' any more?" he says.

The matter is not mere pedantry: getting the pattern right should help researchers to answer meaningful biological questions

— which organisms, if any, forage optimally, and why. Yet for Sims, the qualifier 'like' is not without its uses. It could be useful in probing the complex, interacting factors that affect movement patterns. "Animals often undertake other behaviours interspersed with searching, such as social interactions or predator avoidance," he says, which may weaken the Lévy signal.

Man in the mirror

However the debate plays out, analyses of data from one particular animal, humans, are likely to be increasingly important. Over the past decade, technology has transformed researchers' ability to gather quantitative data on human activities, ranging from patterns of e-mail use to consumers' buying habits. People happily carry radio trackers and tags around in the form of mobile phones. "We finally have objective measurements of what people do," says Albert-László Barabási, a researcher studying human dynamics in this way at the Center for Complex Network Research, based at Northeastern University in Boston. "Our observations don't influence them."

This work can be viewed, perhaps, as the beginnings of a natural ecology of human behaviour, for which understanding patterns of physical movement — the crude equivalent of animal foraging — would offer an obvious first goal. Two years ago, physicist Dirk Brockmann of the Max Planck Institute in Göttingen, Germany, took an indirect stab at the issue using the website www.wheres-george.com, which facilitates the tracking of dollar notes moving through the United States. People can go to the site and enter the date, their location and the serial numbers of dollar bills in their possession. As the bills move, the site shows their changing locations.

Almost 60% of bills starting in New York City were reported 2 weeks later still within 10 kilometres of their starting point. But another 7% had jumped to distances beyond 800 kilometres. If this seems similar to the Lévy pattern, it is. The researchers found that the distribution of distances travelled over a short time follows a power law with a γ equal to about 1.6 (ref. 7).

These data don't directly say anything about the human movements that transport dollar bills. But a team led by Barabási has now gone one step further, using anonymized

mobile-phone data to track the movements of more than 100,000 people over a 6-month period. The statistics, they found, again show the Lévy pattern, although with some additional complexity⁴.

The team found, overall, that the distribution of the distance moved between two subsequent phone calls follows a power law with an exponential cut-off. The best way to explain this pattern, the researchers argue, is through a combination of two effects — first, a real tendency for individuals to move in a Lévy-like pattern, with many short movements and less frequent long excursions, but also a difference between people in the overall scale on which they move, with some people being inherently longer travellers than others. When the researchers normalized the measurements so that the person-to-person scale factor no longer played a part, the data for all the participants fell onto a single curve. "There are a lot of details that make us different," says Barabási, "but behind it all there's a universal pattern."

And what of the albatrosses? Are they an oddity — an error that nevertheless served as the basis for insights into truth? Perhaps. "I think of it like the Bohr model of the atom," says Eugene Stanley, a physicist from Boston University who was one of the original authors. "It was wrong, yet it turned out to be fruitful. The remarkable fact is that flawed data led to a fascinating idea: a general law governing animal movement."

Or perhaps albatrosses do roam the high seas in the way that Lévy might have anticipated, and we will know that in time with better data and analyses. As Viswanathan points out, he and his colleagues' 1999 paper showing the theoretical optimality of Lévy-style foraging provides a good *a priori*

reason to expect that some animals, and quite possibly albatrosses, might exploit this trick.

"Given the power of natural selection," says Viswanathan, "it seems unlikely to me that Lévy walks wouldn't exist somewhere in animal biology. It would be as strange as if vision had never evolved."

Mark Buchanan is author of *The Social Atom*.

"It's well known that log-log axes tend to make relationships look straight."

— Andrew Edwards

1. Viswanathan, G. M. *et al.* *Nature* **381**, 413–415 (1996).
2. Viswanathan, G. M. *et al.* *Nature* **401**, 911–914 (1999).
3. Viswanathan, G. M., Raposob, E. P. & da Luzdoi, M. G. E. *Phys. Life Rev.* (in the press) doi:10.1016/j.plrev.2008.03.002
4. González, M. C., Hidalgo, C. A. & Barabási, A.-L. *Nature* **453**, 779–782 (2008).
5. Edwards, A. M. *et al.* *Nature* **449**, 1044–1048 (2007).
6. Sims, D. W. *et al.* *Nature* **451**, 1098–1102 (2008).
7. Brockmann, D., Hufnagel, L. & Geisel, T. *Nature* **439**, 462–465 (2006).

See Editorial, page 698, and News Feature, page 717.

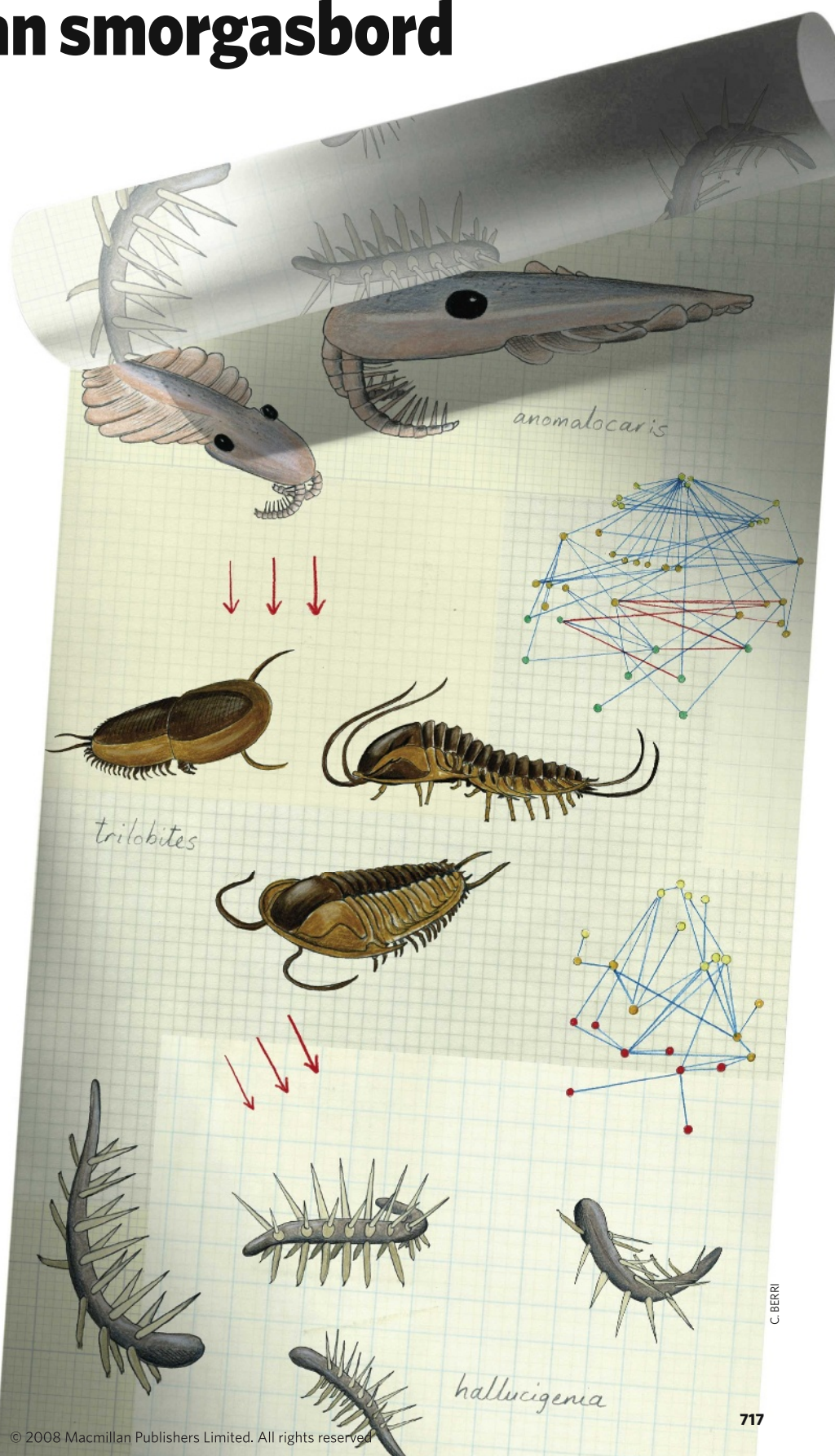
The Cambrian smorgasbord

It began, appropriately enough for research into food webs, over lunch. And appropriately for ambitious and interdisciplinary research, that lunch in 2001 was at the Santa Fe Institute in New Mexico, an organization famous for bringing together creative assemblages of scientists from different fields. Jennifer Dunne, an ecologist at the institute, had been giving a talk about untangling food webs — the networks of who eats who within an ecosystem — by using sophisticated new models. In the audience, and now at her lunch table, was Douglas Erwin, a palaeobiologist based at the Smithsonian Institution in Washington DC. Erwin is, among other things, the curator of the institution's Burgess Shale collection, which despite being more than 500 million years old is one of the best-preserved fossil assemblages in the world.

"We immediately began to brainstorm about whether network analyses could be done with ancient communities," says Dunne. In the past, researchers have assumed that fossil data were not good enough to construct food webs that would be accurate and useful down to the species level; but Dunne's approach, the two thought, might change that. "We weren't sure how the palaeontology community would perceive it," Erwin says. "But we were all pretty excited by the idea." They hoped that it would not merely provide a new look at an ancient community, but that ancient ecologies might cast new light on modern ones, too. Ecologists have shown that modern food webs follow certain rules, such as the distribution of links increasing proportionally with the number of species, or with diversity. But are there universal laws, or are these rules merely a reflection of the world as it happens to be at the moment?

Seeking universals

Ecosystems on other planets would be a great help in sorting out the necessary from the contingent, but remain stubbornly undiscovered. Mimicry in the lab can't capture the necessary subtlety. Food webs from the fossil record might thus be the closest to something completely different that contemporary researchers can ever hope to get their hands on. And the Burgess Shale fauna from British Columbia, Canada, were the obvious choice. Another *lagerstätte* ('resting place' — the term for a site of exceptional preservation), the early Cambrian Chengjiang in southern China, was chosen to serve as a comparison. "One major plus of the Cambrian data is that there is excellent soft-body preservation across taxa," says Dunne. "Most fossil assemblages don't



have that.” Another plus is that the Cambrian is the period in which predators and prey are first unequivocally present in the fossil record; it’s hard to see how there could be interesting food webs any earlier.

In the years after that original lunch, Dunne and Erwin mined existing databases and compiled, reviewed and revised the two Cambrian food webs. “Our first surprise was that we could put these data together,” says Dunne. With the food webs ready, the team then analysed them according to the ‘niche model’ first promulgated by Richard Williams and Neo Martinez¹.

In the niche model, each predator is constrained to eating from one ‘dimension’, and is expected to eat everything in that dimension. So if a dolphin eats tuna and sardines, it is presumed to eat everything between the two on that dimension. In practice, such a dimension maps closely to body size, but that’s not how it is defined; it is a statistical creation that represents many traits, of which body size simply happens to be a significant one in most systems.

Same old same old

Just three of the 17 properties of the Burgess Shale web that they measured fell outside the predictions of the niche model. “Most palaeobiologists would assume that these half-billion-year-old food webs should look really different from modern food webs,” says Dunne. “But it looks like in most ways, at least that we characterize, the organization looks really similar.”²

Dunne feels that the overall success suggests that the niche model is capturing something fundamental: the energetics of the system. Food webs are resource distribution networks, and body size or foraging strategies are limited by the energy expended hunting food and the value derived from eating it. “The patterns that we see may reflect not evolutionary history, but the fundamental physics of the system,” says Dunne.

“The fact that the patterns seem to be robust and they stand up quite well is very impressive,” says Richard Bambach, a palaeoecologist at the Smithsonian.

The clearest difference between then and now is that both Cambrian webs have a higher variability in the number of links per species, something that seems to reflect the higher proportion of predators found in Cambrian fossil assemblages across the world. What this might mean is unclear. Dunne says she thinks there may have been more predators than would be expected today because natural selection hadn’t yet weeded out a category of ‘hyper-vulnerable’ species that modern webs do not have. Conversely, Bambach speculates that the oddities of structure might reflect predators that simply weren’t yet very good, with the best of them not good enough to outcompete the rest.

More webs might clarify things. “It would be interesting to do this throughout the fossil record,” says Bambach. He has mapped changes in animal life strategies, such as the proportion of motile and non-motile organisms³, over geological time and his research suggests that only great extinction events — of which there have been a handful since the Cambrian — are strong enough to break down ecosystems to the point where they reassemble along different lines. That idea may now be testable.

“The general conclusion that basic food-web topology of benthic-dominated marine communities might have been established by the early Cambrian seems to be well supported,” agrees Peter Roopnarine, a palaeobiologist at the California Academy of Sciences in San Francisco who has made less well-resolved ecological models of fossil assemblages. But he is not convinced that the apparent similarities between the ancient and modern ecosystems reflect absolute constraints such as those of energy flow. “There could alternatively be fundamental pathways of web assembly that are favoured over others, either by selection at the species level, or selection at the food-web level itself.” Roopnarine says that his reconstructions hint at long-term community-level selection based on species’ dietary ranges⁴.

But not all palaeobiologists think the new research has much to offer. Nick Butterfield of the University of Cambridge, UK, is unconvinced by the Cambrian webs. The Lobopodian sponge-feeder *Aysheaia pedunculata*, for example, was a centipede-shaped organism a few centimetres long that resembled today’s velvet worms. “Nine specimens of *Aysheaia* are known to be associated with sponges in the Burgess Shale so the authors say all Lobopodia eat all possible sponges,” says Butterfield. “That assumption alone gives them more than 30 connections.” The evidence simply doesn’t go that far, he says.

To get a grip on such problems, Dunne and her colleagues give each link a certainty value. If fossilized gut contents revealed what the organism had been eating, the link was described as ‘certain’ — but only 5% of links, at best, meet that standard. When physical association and morphology indicated a likely interaction, the link was ‘probable’; these links comprised roughly a quarter or a half of the total for the Chengjiang and Burgess webs, respectively. When the evidence was mainly morphological and from what their presumed modern-day relatives do, a link was ‘possible’, as in the case of *Aysheaia*.

The team estimated their model’s sensitivity by removing links, sometimes doing so at random and sometimes according to their likelihood. “In most cases, the results were robust even to the point of pulling out most of the uncertain links,” says Dunne. Roopnarine agrees that the technique looks good: “I think it will be adopted broadly as a measure of variance in web reconstruction.”

Another method of taking uncertain links into account was working with ‘trophic species’ rather than true species. A trophic species is defined as a set of organisms that all eat the same thing, and will typically consist of a

number of true species. *Aysheaia* and its apparently close relative *Hallucigenia* thus collapse into a single trophic species because they both are thought to have fed on sponges, and were in turn fed on by the same predators. Grouping species sacrifices resolution, but can clarify the core functional feeding relationships in a web.

To Butterfield, though, this is folly compounded on folly. “There is no direct evidence that *Aysheaia* ate sponges; and to assert this and then expand it to include all Lobopodia and all sponges over a period of 10 or 20 million years is laughably absurd,” he says.

Butterfield suspects that the webs’ aptness to niche analysis comes from a circularity in the argument. “This claim that all big things eat all little things instantaneously builds the ecosystem,” he says. “They have built the conclusion into their analysis by their assumptions of who’s eating who. And that’s what we want to know.”

Dunne counters by saying that the niche model does not simply assume that big things eat little things; the statistical ‘dimension’ in the niche is not reducible to a physical dimension. “Insects feed on trees; carnivorous pack animals feed on herbivores that are bigger than them; and parasites feed on much bigger hosts.” Would the researchers have identified different trophic relationships if they had been scuba diving in the seas over what would one day be Canada 505 million years ago? “Absolutely,” says Erwin. “But would there be enough difference between that and the web we have now to make a difference? Probably not.” ■

Arran Froom is a UK-based science writer.

- Williams, R. J. & Martinez, N. D. *Nature* **404**, 180–183 (2000).
- Dunne, J. A., Williams, R. J., Martinez, N. D., Wood, R. A. & Erwin, D. H. *PLoS Biol.* **6**, e102 (2008).
- Bambach, R. K., Knoll, A. H. & Sepkoski Jr, J. J. *Proc. Natl Acad. Sci. USA* **99**, 6854–6859 (2002).
- Roopnarine, P. D., Angielczyk, K. D., Wang, S. C. & Hertog, R. *Proc. R. Soc. B* **274**, 2077–2086 (2007).

See News Feature, page 714.



“Our first surprise was that we could put these data together.”
— Jennifer Dunne

J. A. DUNNE

CORRESPONDENCE

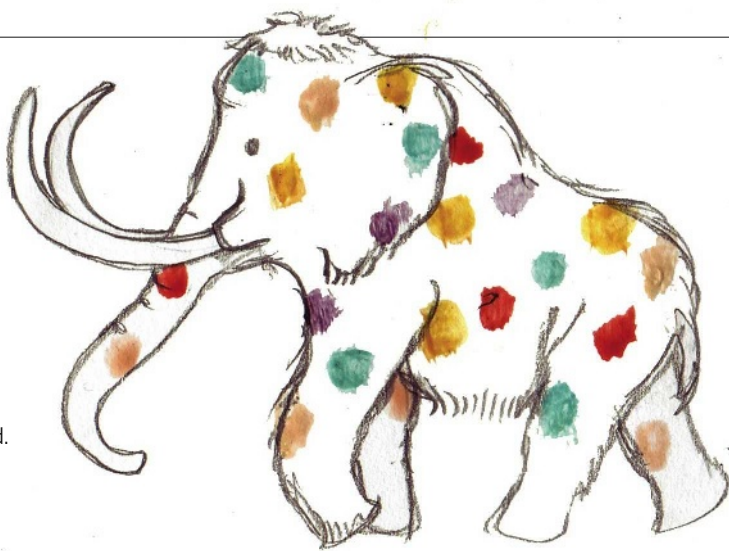
Evolutionary theory: it's on the school syllabus in Mexico

SIR — As Andrew Moore writes in his Commentary 'Science teaching must evolve' (*Nature* **453**, 31–32; 2008), online resources may help to improve the teaching of evolutionary theory in European secondary schools. A broader comparison with other countries is warranted.

In Mexico, there is no creationist movement and the teaching of evolution is encouraged. The Universidad Nacional Autónoma de México has developed high-school programmes based around sequence comparison and phylogenetic reconstruction techniques, as well as the origin of life, early cell evolution and evo-devo issues. There are good reasons for this. First, as every practising scientist knows, understanding evolutionary processes is enriched by an eclectic attitude towards traditional palaeontology and comparative anatomy. Second, an evolutionary perspective is valuable as a teaching tool that provides a unifying framework for bringing together apparently disparate fields from the life sciences.

It is difficult to accept Moore's implication that the origin of life has no place in the evolution syllabus. If not there, where? There are risks in leaving this issue unattended, as shown by the infamous 2004 statement by Dover High School in Pennsylvania that "Intelligent Design is an explanation of the origin of life that differs from Darwin's view ... The school leaves the discussion of Origins of Life to individual students and their families."

Equally disturbing is Moore's questioning of the scientific maturity of origin-of-life studies, which he describes as "far from sound evolutionary theory". How life appeared is not known, but a better understanding of the prebiotic environment and the



recognition of the RNA world as an early evolutionary stage have provided important insights. As molecular phylogeny cannot be extended to an evolutionary period before the emergence of ribosome-mediated protein biosynthesis, it cannot provide direct information about the first living systems.

The inclusion of evolutionary theory in school curricula should teach students how to find the right answers and how to pose the proper questions.

Antonio Lazcano, Arturo Becerra
Facultad de Ciencias, UNAM, Ap.
Postal 70-407, Cd. Universitaria,
04510 Mexico, D.F., México
Juli Peretó Institut Cavanilles de
Biodiversitat i Biologia Evolutiva,
Universitat de València, Ap. Postal
22085, E-46071, València, Spain

Evolutionary theory: don't skimp on teaching its history

SIR — In his Commentary 'Science teaching must evolve' (*Nature* **453**, 31–32; 2008), Andrew Moore criticizes the absence of the past four decades' efforts in gene-sequencing technology and bioinformatics from European secondary-school curricula. He notes that "phylogeny based on similarity of form is fundamentally unsound because of the adaptation and convergent evolution witnessed in nature". There is

self-evident truth in this, but molecular phylogenetics is also based on similarity of form, albeit DNA's. Many of the computer programs used to seek out molecular relationships among organisms are used to determine morphological relationships — thereby revealing those many instances of convergence.

Data are data, whether they be molecular or morphological. If students are short-changed when it comes to the teaching of evolution, it is in its history, which is highly relevant today.

David M. Williams Department of
Botany, the Natural History Museum,
Cromwell Road, London SW7 5BD, UK
Malte C. Ebach International Institute
for Species Exploration, School of Life
Sciences, Arizona State University, PO
Box 874501, Tempe, Arizona 85287-
4501, USA

Learning to navigate the geography of citation indexes

SIR — In his Correspondence 'Hall and Keynes join Arbor in the citation indexes' (*Nature* **452**, 282; 2008), Daniel Postellon describes the distinguished careers of Milton Keynes, Walton Hall and Ann Arbor. In the last case, I note that Professor Arbor has an h-index of 1 from the Web of Science database provided by Thomson Scientific's ISI Web of Knowledge. This is based on her five citations for the

year 2007: two articles, two letters and one abstract.

An author search in the Web of Science reveals that Chevy Chase (not to be confused with Chevy Chase, Maryland) has co-authored a letter with Howard Kaplan (*H. Kaplan Am. Sci.* **96**, 3; 2008). My own institution, Wilfrid Laurier University, is also an author (*S. Cadell et al. J. Palliat. Care* **23**, 273–279; 2007).

Irrespective of how these errors are created, the rising use of systems in which citation information moves directly from the search of a database or citation index to a bibliographic management system, and then into a reference list, means that inexperienced students and researchers who are not savvy enough to detect these errors will propagate them further.

Vigilance is required by all users of citation indexes and databases.

Debbie Chaves Wilfrid Laurier
University, 75 University Avenue West,
Waterloo, Ontario N2L 3C5, Canada

Vaccine failure is not a 'crisis' for HIV research

SIR — In your Editorial 'Broken promises' (*Nature* **452**, 503; 2008), you say: "Decisions to move Merck's vaccine candidate and a previous failed candidate into clinical trials were based only partly on science." I would like to clarify the position on the Merck vaccine trial, as there is no evidence to support this assertion.

The scientific consensus was that the phase IIb efficacy trial of this vaccine candidate was justified. Inducing CD8 T-cell responses in the majority of vaccine recipients by strategies other than live attenuated vaccines has proved extremely difficult, and the adenovirus serotype-5 platform was the first to overcome the problem.

The trial was designed to be smaller than a traditional phase III and included an interim analysis after the first 30 infections to

ensure participant safety. It was this analysis by the Data Safety Monitoring Board that sparked the concern about safety with the adenovirus vector. The trial has certainly raised questions about whether the induction of HIV-specific T-cell responses is a viable vaccine strategy, but these questions are best addressed by analysing the data.

Particularly important are questions regarding the effect of the quality, quantity and specificity of the vaccine-induced CD8 T-cell response on post-infection viral load control. These questions could not be addressed without a vaccine approach that actually induced CD8 T-cell responses in most recipients.

The Merck vaccine was the first such candidate, so it is a misleading exaggeration to claim that its failure is a "crisis" for HIV vaccine research. A journal with *Nature's* long history is well placed to know how likely first-time successes are in science.

In my view, your Editorial risks reinforcing the unrealistic expectations of science that you erroneously imply were promoted by the HIV-vaccine researchers involved in the Merck trial.

Richard Jefferys Michael Palm Basic Science, Vaccines & Prevention Project, Treatment Action Group, 611 Broadway, Suite 308, New York, New York 100012, USA

Fixing hiring practices means asking the right question

SIR — Whenever Spain's research or academic life is analysed (see, for example, *Nature* **451**, 1029; 2008), a comment invariably follows on hiring practices and regulations (*Nature* **453**, 26–27; 2008). This obsessive repetition points to a deep problem in that country, but these issues are shared by many other nations and so deserve international attention.

The question that is usually asked or answered concerns 'how': how should the academic system

be regulated to avoid widespread poor practices in hiring? As an academic, I would start by asking 'why': why is it that the hiring decisions of academics in some countries cannot be trusted and apparently need to be enforced by tight regulations, whereas this is not the case in other countries?

Remedies don't work well without a good diagnostic.

Emilio Artacho Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Germline modification carries risk of major social harm

SIR — Your Editorial 'New sources of sex cells' (*Nature* **452**, 913; 2008), on the potential use of pluripotent stem-cell-derived gametes (PSCDGs) for germline genetic modification and enhancement, suggests that the prospect of stem-cell-derived gametes could trigger renewed calls for regulating human biotechnologies. In those discussions we must, as you warn, be wary of impeding basic research. But we must be equally willing to draw lines proscribing socially pernicious applications.

Germline (that is, inheritable) modification is the most socially consequential and ethically dubious application of human biotechnology; its implications have been explored from a wide range of perspectives. Most of these discussions have focused on the social meaning and repercussions of genetic manipulation of the human species, not on the moral status of human embryos. One can strongly support human-embryo research and still oppose germline modification. Conversely, some opponents of abortion rights support germline modification, so long as no embryos are destroyed in the process.

Although there may be scientific and therapeutic benefits from research on PSCDGs, the case for

any such benefits from using these cells for human germline modification is weak, whereas the likelihood of substantial harm is great. One consequence could be inequalities between those who can afford genetic enhancements and the majority who cannot. Human-germline modification could lead to the emergence of 'genetic castes', creating vast social rifts, with horrific consequences. Such considerations have already prompted many countries — including most of Europe, Canada, Japan, South Africa and Brazil — to prohibit germline modification.

Marcy Darnovsky Center for Genetics and Society, 436 14th Street, Suite 700, Oakland, California 94612, USA

Global database is needed to support adaptation science

SIR — Cynthia Rosenzweig and colleagues have taken a critical step towards a global synthesis of biological and physical impacts attributable to climate change (*Nature* **453**, 353–357; 2008). They have expanded the database used in the recent Fourth Assessment from the Intergovernmental Panel on Climate Change (IPCC). We feel there are still many more studies that could be compiled to improve geographical and ecosystem imbalances and to provide a more comprehensive overview. Given the resources and effort required to identify these, it is beyond the scope of a limited number of individuals.

Climate-change impacts are being reported in a burgeoning literature from every ocean and continent. It is now time to shift the emphasis from proving climate impacts to providing key support for adaptation science. This requires a publicly accessible global database to collate research into climate-change impacts research and to allow the scientific community and the IPCC to focus additional efforts on attribution and adaptation.

Global cooperation and sharing model outputs have led to breakthroughs in understanding the climate system, as illustrated by the scientific community's assembling of the physical-science basis (IPCC working group I). We suggest the impacts community (working group II) should follow a similar path.

Just as the IPCC endorsed the Program for Climate Model Diagnosis and Intercomparison data repository for model projections, so could it endorse an impacts repository, with researchers able continuously to upload research results that can then be quality-assured by a verification panel. The panel could be selected on the basis of expertise across a range of disciplines, and the repository website hosted by an international body, such as the International Geosphere-Biosphere Programme. Similar global initiatives underline the value of such an approach. These include the Ocean Biogeographic Information System (set up by the Census of Marine Life) and GenBank (National Center for Biotechnology).

This approach would allow greater efficiency, transparency and completeness in the compilation process, facilitate rapid identification of knowledge gaps and allow broader, expert-driven quality control of the interpretation of biological, aquatic and terrestrial data. This pathway to transparency and rigorous global synthesis is critical for a potential Fifth Assessment Report and for a world relying on scientific guidance.

Elvira S. Poloczanska, **Alistair J. Hobday** Climate Adaptations Flagship, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia
Anthony J. Richardson Climate Adaptations Flagship, CSIRO, Marine and Atmospheric Research, PO Box 120, Cleveland, Queensland 4163, Australia and University of Queensland, Department of Mathematics, St Lucia, Queensland 4072, Australia

BOOKS & ARTS

Command and control

A biography of botanist Joseph Dalton Hooker illustrates how science switched in the nineteenth century from being a hobby of aristocrats to a profession paid for by governments.

Imperial Nature: Joseph Hooker and the Practices of Victorian Science

by Jim Endersby

University of Chicago Press: 2008. 400 pp.
\$35, £18

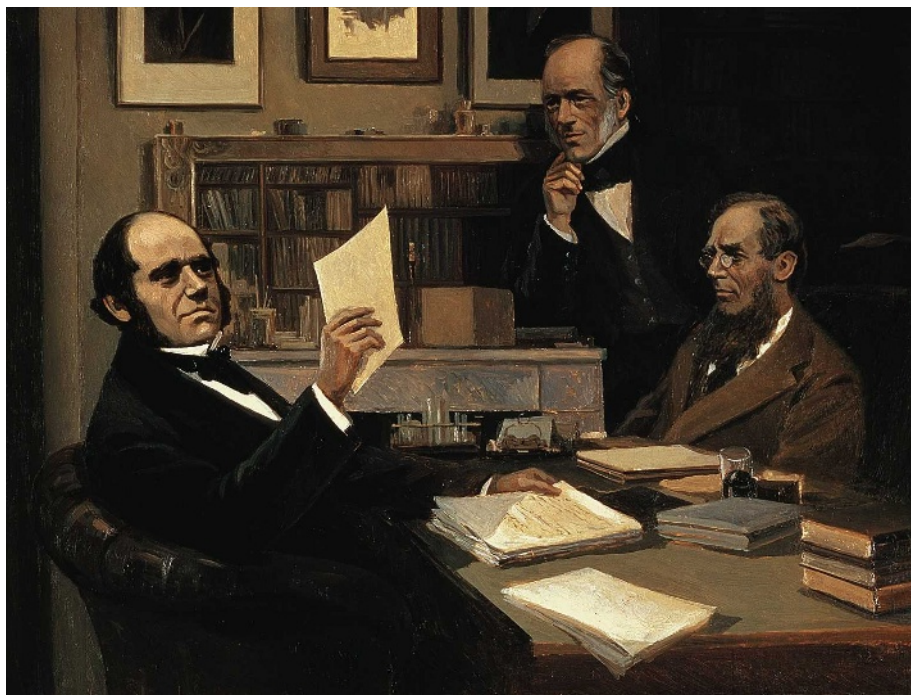
As we approach next year's frenzy of celebrations for Charles Darwin's bicentenary and the 150th anniversary of his publication *On the Origin of Species*, it is important to remember other naturalists who worked in the mid-nineteenth century. *Imperial Nature* chronicles Joseph Dalton Hooker, who transformed a royal pleasure park into the scientific institution now called the Royal Botanic Gardens at Kew in London.

Science, particularly natural history, switched at this time from an activity practised by aristocrats to one paid for by governments. Jim Endersby provides a refreshing record of how scientists worked during this transition, rather than an analysis of the theories they generated. His contention, with which I agree, is that the practice of science provides the context necessary for understanding how theories advanced; without this background, scientific progress looks too simple, and leaps seem extraordinary.

Hooker was a close friend of Darwin. At the Linnean Society of London, he helped to engineer the joint reading of Darwin's abstract of *On the Origin of Species* alongside Alfred Russel Wallace's paper on natural selection, which Wallace had sent from the field in southeast Asia. Like Darwin, Hooker is usually portrayed as a whiskered gentleman of the establishment. Unlike the independently wealthy evolutionist, Hooker had to earn his living from science. Endersby describes Hooker's desire to join the scientific elite and practise 'philosophical' rather than 'paid' science. Returning from an expedition to Antarctica on the ship *Erebus*, Hooker wrote to his father: "My hope and most earnest wish is to be able to on my return home devote my time solely to botany."

Hooker epitomizes the advent of the professional scientist, but he was at pains to give the impression that he worked purely for the love of science, not for pay. Perhaps the relatively low salaries in some modern fields are a legacy of the ambivalent attitude of Victorian scientists such as Hooker towards remuneration.

Hooker managed Kew Gardens, which was



Joseph Hooker (right) and Charles Lyell (standing) discussed evolutionary theory with Darwin (left).

government funded and part of the UK civil service, as a private fiefdom. In the early 1870s, this style caused a stand-off between Hooker and Acton Smee Ayrton, essentially head of the civil service in prime minister William Gladstone's government. Ayrton attempted to impose civil-service hiring and procurement rules at Kew, but Hooker fought against it. Hooker's friends defended him in letters published in *Nature*, describing him as a selfless man working for the greater good, who was owed a living by the nation.

Hooker backed down and was forced to apologize to Ayrton for insinuating he was a liar. Gladstone commented on Hooker's behaviour, observing that "scientific men ... have a great susceptibility" and are "not accustomed to enter in our sturdy conflicts". A scientist giving evidence to parliaments today might agree.

In trying to join the scientific elite, Hooker had the additional burden of being a botanist. Botany and natural history were then low-status disciplines, below the physical sciences, chemistry and geology, but above medicine. Botany was not taken seriously because it was accessible to anyone, including

women, and was not considered intellectually demanding. It was also associated with gardening and horticulture. Hooker realized that to make his name, he must invent a 'philosophical botany', a science of plants that laid out general rules rather than describing details. But he needed more data than one man could generate — he needed an army of collectors.

Hooker's relationships with these collectors are the most fascinating part of Endersby's book. As a young man, Hooker collected plants in Antarctica and India, and some of the men he met on his travels remained correspondents and collectors for Kew throughout his career. Hooker rarely paid for material sent to Kew but operated a barter system. He exchanged books and equipment for plants from New Zealand and Australia, ensuring that the collection of dried plants at Kew became global. Collectors, however, occasionally showed independence. Hooker reprimanded those who dared to describe new plant species from their area — species could be described only at Kew, after comparison with its collections. That this was accepted seems extraordinary, but the collectors

EVSTAFIEFF; DOWN HOUSE, KENT, UK/BRIDGEMAN ART LIBRARY

NATURAL HIST. MUS., LONDON

needed Hooker and his gifts: they earned status by being involved in the new science.

Hooker had a strong vested interest in being the sole person to define a species. Like Darwin, he deplored 'species-mongers', today known as splitters, who described variants as species in their own right. Hooker's theories of plant distribution depended on a broad species concept, so it was important that he maintained control of the definitions. He also defined localities. Collectors of the time recorded broad regions of sample origin, such as southern India, rather than specific



Hooker's drawing of a red alga, *Delesseria*.

coordinates. Today we do the reverse using georeferencing.

According to historic accounts, Hooker was a reluctant convert to the theory of evolution by natural selection. Enderby shows that the story was more complex. Hooker supported Darwin but did not think that evolutionary theory affected botany in practice, noting that the evolutionist must "employ the methods and follow the same principles that guide the believer in their being actual creations". This view sounds familiar to anyone who was involved in the pattern-cladistics furore of the 1980s, when a set of systematists

were accused of being 'anti-evolution'. Evolution by natural selection remains the most robust explanation for the generation of biological diversity. Study of what that diversity entails can be theory-free, as Hooker contended, but studying diversity in the light of evolution is more satisfying.

It is surprising how familiar the debates of nineteenth-century science sound today. By concentrating on practice, *Imperial Nature* reminds us that although theories are important, the evidence on which they are based comes from many sources and through many cultures. One hopes that Hooker's attempt at central control could never happen today, with our vibrant, diverse and more equitable communities.

Sandra Knapp is a plant taxonomist in the Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Staving off the global food crisis

The End of Food

by Paul Roberts

Houghton Mifflin/Bloomsbury: 2008.
416 pp. \$26/£12.99

Sometimes an author gets lucky, or is truly prescient. He can work for years researching a complex and obscure topic, only to see it hit the headlines just as his book is published. Suddenly, the topic is hot.

Food is hot. If high supermarket prices have not grabbed the average citizen's attention, the world food crisis surely has. With food riots from Haiti to Egypt and panic-buying of rice in Hong Kong and Vietnam, food scarcity is the topic of the day. Following on from his earlier best-selling book *The End of Oil*, Paul Roberts's *The End of Food* taps into these timely concerns.

Food crises tend to recur in history. The most severe in recent times was the world food crisis of 1973–75. Even the Old Testament of the Bible talks of years of glut and famine, and the role of good governance in smoothing out supply.

Are our worries about food different this time? Perhaps in the future we will see constantly high prices, the re-establishment of food scarcity in the developed

world after decades of surplus, and widespread hunger. Or perhaps a technological solution will lessen the tension between a growing human population and the natural resources that feed it. Will there be a continuation of the trends that Roberts documents so well, of perpetually lower prices, greater reliance on world trade to source the cheapest commodities, the spread of meat-intensive diets with increasing affluence, and more land used to grow corn for ethanol to fuel our cars?

Roberts's answers are clear. The global food system, as it is currently structured and driven, is heading for a cataclysm. Roberts offers a sobering scenario of a 'meltdown': "We are already growing fatter (and hungrier), depleting more soil organic matter, drawing down more water tables, using more fertilizers and pesticides, losing more acres of forests and farmland." Consequently, he warns, "There is no longer the possibility of discrete failure; a collapse of one part of the system will have extraordinary ramifications for everyone else."

The End of Food makes the case that system-wide collapse is inevitable. Roberts starts by recognizing that economic forces drive the world food system, although our basic biological needs for nutrition have not changed since we evolved. This tension between food as an economic commodity — produced, processed, even speculated on as if it were copper or steel — and as a biological necessity is not new. But Roberts argues that globalization of our food supply and the westernization of dietary demand have driven the entire system irrevocably out of balance.

The result is a list of woes. The industrialization of the food industry creates a need for sources of cheap inputs and continual supply of new products. The retail revolution has led to a tendency to offer 'supersize' portions to push up demand. Obesity is the



Supporters of the Gabriela political party protest against rising food prices near president Gloria Macapagal-Arroyo's palace in the Philippines in April.

LIUWANG/AFP/GETTY IMAGES

consequence of these two transformations of the food system. Global trade is able to supply progressively cheaper food, at high cost to humans and the environment, yet there is a paradox of plenty amidst widespread hunger. Food-borne diseases resulting from modern farming techniques for livestock have also sharply raised the probability of an uncontrollable pandemic.

Roberts is not hopeful of a solution to these problems because of the economic forces that dominate. He avoids conspiracy theories, but distrusts the coordination between the producers and consumers that is central to the capitalist system. He is not alone. To paraphrase British prime minister Winston Churchill on democracy, capitalism is the worst way to organize society's economic activities, except for all of the alternatives. Communist Cuba, for example, returned to a system of local food production using human and animal power to produce a nutritionally adequate diet. Roberts accepts this is hardly a good global solution.

There are two approaches to making our food system safer and more sustainable, yet still accessible to the world's population, which is expected to increase by two billion during the next 30–50 years. Both strategies should be pursued simultaneously.

First, and ironically in view of Roberts's critique, the global capitalist system needs to be harnessed to help solve problems of food scarcity, pricing and inequity. Fortunately, it is already moving in the right direction. The rising cost of energy makes many elements of the food industry unprofitable. High fertilizer prices, high transportation costs and high meat prices all push the system towards less intensive, locally produced and healthier alternatives.

Second, good public policy and government investments in food and agricultural research can make a big difference. More-effective regulation, better-educated consumers and healthier school environments can all follow from elected officials who care about the quality and quantity of the food they and their families eat.

The End of Food is a call to arms. But there will be no revolution. We will all be buried in our oversized coffins before a radical solution comes, because change will be gradual. So we must get on and fix the system we have.

C. Peter Timmer is a visiting professor in the Program on Food Security and the Environment, Stanford University, Stanford, California 94305, USA, and author of *A World Without Agriculture: The Structural Transformation in Historical Perspective*.

Horse power unbridled

The Horse

American Museum of Natural History, New York

Until 4 January 2009.

Noble, speedy, dependable and strong, the horse changed the course of human history. The domestication of *Equus caballus* some 6,000 years ago supplied people with a reliable year-round source of food, enabling them to settle in villages and ride to communities farther afield. Horses have since helped humans plough fields, carry supplies, haul water, mine coal, fight battles and conquer continents. The animals raced through the earliest art and have a renowned place in religion and sport.

They also inspired the invention of trousers.

As explained in *The Horse*, an exhibition at the American Museum of Natural History in New York, trousers were designed for the comfort of riders, and only later adapted for more general use.

They appear on a fifth-century BC urn from southern Italy that depicts a duel between an Amazon woman and a Greek warrior. Mounted on a horse and clothed in striped Persian pantaloons, the Amazon aims her spear at the Greek, who fights back helplessly on foot. The exhibition emphasizes that humans owe more to horses than practical fashion; curator Ross MacPhee calls the animals "the first multipurpose engines, animated machines".

Horses helped power the Industrial Revolution: they hauled goods from steamships or trains to warehouses and markets. Horses often worked in tandem with machines; by the 1870s, more than 300 US patents had been issued for horse-powered machinery. One horse-drawn contraption on display is a three-metre-long, three-tonne steam fire engine made of iron and steel, dating to 1896. Before the mid-nineteenth century, men pulled and hand-pumped fire engines. A draft horse is strong enough to lift 150 kilograms a distance

of 30 metres in one minute, or, as Scottish inventor James Watt calculated in the 1770s, 33,000 'foot-pounds' per minute, a quantity he defined as one horsepower. Museum visitors can pull down a lever and measure their own strength: after some practice and puffing, and by using both hands, I scored about a fifth of one horsepower.

The great mobility of horses derives in part from their ability to graze and digest grass on the go, enabling them to walk all day. Their speed hails partly from their single hooves, which evolved around 12 million years ago in a genus called *Pliohippus*, relatives of modern-day *Equus* — single hooves enabled them to run farther and faster than their three-toed, forest-dwelling ancestors.

These adaptations helped the cavalries of Genghis Khan and his descendants to conquer most of Asia, the Middle East and Russia. Mongol horsemen

did more than pillage: they spread innovations along the Silk Road such as the bow fiddle, which originated in central Asia around 1,000 years ago and had a bow strung with horsetail hair.

A modern Mongolian example of the *morin khuur*, the horsehead fiddle, is on display, as are other forms of horse-inspired art: reproductions of 32,000-year-old paintings of horses

galloping gracefully along the walls of the Chauvet cave in France; children's toy horses; and a garland-trimmed, 2-metre-long terracotta horse from Tamil Nadu, India, created as an offering to local gods.

A powerful bond exists between humans and horses. A film exhibit shows an organization called GALLOP, based in Brooklyn, New York, which provides therapeutic horseback rides for people with disabilities. In one segment, Sarah Lisker, a girl who has cerebral palsy, is lifted from her wheelchair onto a horse and parades around confidently, assisted by an instructor. "What did the horse teach you?" the instructor asks. "That I'm his friend," Sarah replies, and then murmurs, "Trot on."

Josie Glausiusz is a journalist based in New York.



A brass toy horse and soldier from India.

AMNH/D. FINNIN

How the car became king

Fighting Traffic: The Dawn of the Motor Age in the American City

by Peter D. Norton

MIT Press: 2008. 396 pp. \$35, £22.95

Fighting Traffic tells the story of how cars came to dominate our streets and how new roads reshaped our cities, with many parallels to traffic management today. Technology historian Peter Norton's book focuses on the 1920s, when motorists were in the minority. Motor cars, or pleasure cars, were a luxury and not the necessity they have become in many countries. Early motorists fought for the right to use the streets against strong opposition from non-motorists, who called cars 'juggernauts' and 'villains'. Dislike of cars was fuelled by the rise in traffic accidents, which in 1924 claimed more than 20,000 lives in the United States. Marian L. Telford, a safety activist, wrote in the *National Safety News* of October 1926 that in Chicago, Illinois, "nearly half of children struck down in city streets were on their home block, a fact indicating that unsupervised street play was probably a much bigger risk factor than journeys to schools or stores".

The key question at the time was whose street was it? Were streets places for social interaction, including children's play and commerce, or were they avenues for movement? And what kind of movement was to be allowed there — pedestrian passage or vehicular transport in the form of trams or cars? Norton considers the question as one of social justice.

Another issue emerged from the controversies of the time: efficiency. Congestion in the streets grew because of the chaotic interference of pedestrians with traffic. Free access to the streets, without rules, slowed transportation between business areas, causing adverse economic effects. In line with economists' views at the time, support for 'free market' use of the street was tempered by the realization that the 'invisible hand' of the market was not imposing natural order as it should. Regulation was called for to improve safety and lessen congestion. Engineer Ernest Goodrich, addressing the 1923 US conference on city planning, argued that "street capacity can be increased effectively by regulating traffic".

The concept of traffic regulation is not new. Julius Caesar restricted freight to the night hours to ease daytime congestion and allow pedestrians and other traffic to use the streets in imperial Rome. What was new in the 1920s was the growth of scientific management. This approach was inspired by the



Early motorists in 1920s Chicago faced opposition from pedestrians because of the rise in traffic accidents.

nineteenth-century mechanical engineer Frederick Winslow Taylor, who promoted the idea that systematic analysis was superior to trial-and-error methods. This led to the first traffic survey in Chicago in the 1920s and to the 1924 traffic plan in Los Angeles, California, by landscape architect Frederick Law Olmsted. Regulations introduced progressive measures to separate pedestrian and motor traffic, including signals and parking controls. Today, cities across the world use similar surveys and management schemes.

Regulation also affected non-motorists. Specialized children's playgrounds were introduced, surrounded by fences. Pedestrian movement was controlled by restricting where people could cross streets, along with patrols to help children cross in safety.

By the end of the 1920s, regulation had not cured congestion, and car manufacturers saw the need for more street space to increase car sales. Engineer and future US president Herbert Hoover convened a conference on traffic in 1928, which resulted in an appeal to increase 'floor space' for vehicles. According to Norton, this heralded the triumph of 'motordom' and the reconstruction of the city itself.

A new vision for the city's form was propounded by General Motors at the *Futurama* show at the World's Fair of 1939–40 in New York. Elevated motorways and high-density skyscrapers dominated the new city design.

The dawn of the motor age in the United States was finally achieved in 1955 by President Dwight D. Eisenhower's national programme

of interstate highways. This massive building project provided the freedom of movement that was essential for the economic development of the whole country and improved driving safety. In 1955, it was forecast that the 16 fatalities per 100 million vehicle miles observed in 1938 would be reduced to 5; this was achieved by 1966.

The interstate highway system was probably a major force encouraging the sprawl of US cities. City form always follows the pattern of new transport provision. This was predictable from the shape of Greater London's rapid growth in the late nineteenth century following the introduction of suburban rail networks. As envisaged in the 1940s, internal restructuring of cities also took place, to the great detriment of their historical cores.

It is fascinating to see that the same arguments used in the 1920s to oppose the building of roads are still used. Engineer Arthur S. Tuttle's visionary words in the April 1927 issue of *Automotive Industries* could have been spoken by today's politicians: new roads "would be filled immediately by traffic which is now repressed because of congestion". The argument that shifting traffic away from city centres "threatens downtown commerce and property values" is also widespread today.

Fighting Traffic traces the main issues around the introduction of the motor car. As the author explains, "The case presented in this book is that before the city could be physically reconstructed for the sake of motorists, its streets had to be socially reconstructed

THE ART ARCHIVE/NATIONAL ARCHIVES, WASHINGTON DC

as places where motorists unquestionably belonged.”

My only reservations about this excellent book are its over-reliance on sociological jargon and that there is not enough discussion of the economic benefits that made the car so successful. In most advanced economies, motor vehicles account for more than 70% of passenger and freight mobility. The increase

in mobility has reduced the monopoly power of city-centre landlords, allowing better and cheaper housing in the suburbs. At the same time, the monopoly power of central business has been reduced by the competition offered by out-of-town retailers. These benefits have improved economic efficiency, but at a considerable environmental cost. Maybe the author should consider writing a sequel

about the ‘dusk of the motor age’ as strong environmental, economic, social and political pressures increasingly threaten motorists’ freedom. ■

Marcial Echenique is professor of land use and transport studies at the Department of Architecture, University of Cambridge, 1–5 Scroope Terrace, Cambridge CB2 1PX, UK. He and Andrew Saint edited *Cities for the New Millennium*.

In Retrospect: The chromosome trail

A new translation of Theodor Boveri’s 1914 monograph brings the early origins of contemporary cancer research to a wider readership, contends **Robert A. Weinberg**.

Concerning the Origin of Malignant Tumours

by Theodor Boveri. Translated and annotated by Henry Harris

Cold Spring Harbor Laboratory Press: 2008. 82 pp.

Theodor Boveri (1862–1915) was a towering figure in cell biology and cancer research during the early twentieth century. Trained as a zoologist, he probed the workings of the nucleus and cytoplasm of cells by perturbing them in invertebrate embryos. He was particularly interested in the contributions of chromosomes to cell behaviour. His speculative monograph on the connection between chromosomal anomalies and their role in cancer, published in German in 1914, sealed his scientific reputation.

This monograph, an important source of the origins of contemporary cancer research, has now been translated into English for the second time by Henry Harris, a pioneer in the discovery of tumour-suppressor genes. Boveri’s widow and co-worker, Marcella O’Grady Boveri, published the first translation in 1929. Realizing that German biomedical research had begun to lose its dominance worldwide and that German was no longer the sole language of science, she undertook to bring her husband’s work to the attention of the non-German-speaking scientific world. Harris’s translation presents Boveri’s ideas in more contemporary English and is far more accessible. Frequent explanatory footnotes set the arguments in the context of recent findings and, on occasion, in light of twenty-first-century cancer research.

Boveri is remembered for the seminal idea, laid out in *Concerning the Origin of Malignant Tumours*, that chromosomes are the seats of cell heredity, and that the atypical chromosomes

often seen in cancer cells are the basis of these cells’ aberrant behaviour. In truth, much of the credit for this prescient suggestion should go to the pathologist David von Hansemann, who, beginning in 1890, published a series of papers on the subject. Von Hansemann noted the tight association between abnormal numbers of chromosomes (aneuploidy) and malignant tissues (neoplasias), but he did not conclude that one invariably causes the other. Boveri notes von Hansemann’s contribution in passing and presents evidence for just that conclusion.

To this day, von Hansemann’s work remains relatively unknown. He lacked a doting widow to translate his own body of work. And Boveri was a more dominating personality — an otherwise uncritical student described him as a “vehement, inflexible and relentless assailant”. Boveri’s eclipsing of von Hansemann is reminiscent of how Rudolf Virchow, widely credited with the dictum that all cells arise from the division of pre-existing cells, lifted the idea from the less well-connected Robert Remak, a neurologist who repeatedly reported this finding in the years before Virchow’s famous 1858 paper.

Boveri manipulated sea-urchin eggs and embryos. He stripped eggs of their nuclei or fertilized them with multiple sperm. By so doing, he produced evidence for the theory, already in wide circulation, that the determinants of heredity lie in the nucleus, not in the cytoplasm, and that faulty cell division is responsible for the presence of abnormal numbers of chromosomes. From his later experiments, he deduced that each chromosome within a single nucleus must carry a distinct type of heritable information and that proper embryonic development depends on the inheritance of the correct complement of chromosomes.

Boveri seems to have been unaware of the dramatic revolution in genetics that followed

the rediscovery in 1900 of Gregor Mendel’s all-but-forgotten research into inherited traits. Soon after the rediscovery, several scientists noted the striking parallels between Mendel’s genetic determinants and the behaviour of chromosomes — by then known to be present in pairs in ordinary cells, and only singly in sperm and eggs. None of these insights features in the 1914 monograph.

He also studied the two poles that appear at the opposite ends of a cell that is about to divide. These poles anchor and organize the spindle fibres that pull apart two sets of chromosomes during cell division, ensuring each daughter cell receives one complete set. Boveri’s focus on the spindles that arise in cells with four poles, rather than the usual two, as the main source of abnormal chromosome numbers now seems simplistic. But it was a good start given the limited experimental tools at his disposal.

What makes Boveri’s writing worth reading almost a century later is his ability to distil complex information, including the ideas and observations of others, into brilliant, incisive syntheses. For those biologists among us who are interested in how we got to where we are, this book makes fascinating reading, even if it does not lay down an accurate trail of previous discoveries. Cancer research over the past century has involved a succession of blind alleys and detours, mountains of largely uninterpretable observational data and the occasional brilliant leap forward. This book takes us back to the humble beginnings of this now thriving field. ■

Robert A. Weinberg is a member of the Whitehead Institute for Biomedical Research and professor of biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. He is author of *The Biology of Cancer*.

ESSAY

Talk of the tone

To appreciate how our species makes sense of sound we must study the brain's response to a wide variety of music, languages and musical languages, urges **Aniruddh D. Patel**.

Music engages much of the brain, and coordinates a wide range of processing mechanisms. This naturally invites consideration of how music cognition might relate to other complex cognitive abilities. Language is an obvious candidate, as, like music, it relies on interpreting complex acoustic sequences that unfold in time.

Whether music and language cognition share basic ways of making sense of sound has only recently begun to be studied empirically. An exciting picture is emerging. There are more connections between the domains than might be expected on the basis of dominant theories of musical and linguistic cognition — from sensory mechanisms that encode sound structure to abstract processes involved in integrating words or musical tones into syntactic structures. Comparative music–language research offers a way to explore the processing underlying both domains. Such work may lead to a

deeper understanding than could be achieved by studying each domain in isolation.

Practically all the work in this area, including my own, has focused on Western languages and musical traditions. This has been a productive starting point because both have been studied the most deeply theoretically and empirically. It is now time to broaden the cultural scope of comparative cognitive research.

World view

Fascinating questions about music and language emerge when one looks beyond Western culture. For example, what do musical scales (such as 'do-re-mi-fa-so-la-ti-do') have to do with language? At first glance, the answer seems to be 'very little'. No human language, not even those in which a word's pitch can change its meaning, organizes pitch in terms of musical scales.

In the West there has been a tendency to view the structure of our musical scales as a product of nature, reflecting the laws of acoustics and of auditory physiology. Musical scales and their

constituent pitch intervals are implicitly considered a sort of mathematics made audible, a view that can be traced back to Pythagoras' experiments with vibrating strings. Elsewhere, the Javanese *pelog* and *slendro* scales have pitch intervals not found in Western scales, and there is substantial variation in the tuning of these from one gamelan orchestra to the next. The subtle microtones of Arabic and Indian music, which enrapture native listeners, can sound out of tune to Western ears.

Hence the structure of our Western scales cannot be considered universal. What is universal across cultures is the use of a small and consistent set of pitches and intervals within the octave as a framework for performance and perception. Within any given culture, listeners absorb this system simply through exposure and unconsciously use it to extract discrete pitch categories from signals in which pitch varies continuously (as in song, where there are often smooth glides between notes).

Viewed in this way, there is a conceptual connection to the learning of sound categories in language. Each language has its own set of distinctive speech sounds or phonemes, which native listeners learn implicitly as part of making sense of the sound stream that reaches their ears. Music uses pitch to distinguish the notes and intervals of the scale; language largely uses timbre to distinguish phonemes.

Crucially, both domains rely on the ability of the mind to create and maintain discrete sound categories in the face of complex and time-varying acoustic signals. Speech and music may share some of the basic brain processes for forming sound categories, even though the end products are built from different acoustic 'stuff'.

Beat poetry

Non-Western cultures prompt another question about music–language relations: to what extent are basic aspects of rhythm perception universal? On the basis of research in Western European countries, it has been claimed for more than 100 years that a ubiquitous aspect of rhythm perception is the tendency to hear grouping or phrasing in auditory patterns in a particular way. For example, when events in a sequence vary in duration (such as tones of alternating lengths: ... long–short long–short ...), it is claimed that listeners hear long events as final. In other words, the perception would be of a repeating short–long group,



rather than the logically possible alternative of a repeating long–short group. However, recent research on non-Western rhythm perception shows that even this basic aspect of our interpretation of sound varies between cultures. Many Japanese adults hear sequences of this sort as repeating long–short groups.

This difference is unlikely to be innate, so should be relatable to some characteristic auditory patterns in Japanese culture. One might assume it reflects familiar musical rhythms. Studies by my research team in collaboration with colleagues in Japan suggest instead that the key factor is language. English and many Western European languages put short grammatical words before the longer content word to which they are syntactically bound (for example, ‘the book’, *le livre*, *het boek*), creating an inventory of short–long linguistic patterns. Japanese puts grammatical words after their associated content word (for example, *hon-wo*, where *hon* means ‘book’ and *wo* is a grammatical particle), creating frequent long–short acoustic chunks in the language.

I think that language syntax strongly influences a listener’s ambient rhythmic environment and shapes how they hear even non-linguistic patterns at a basic level. We are now pursuing this hypothesis in further cross-cultural and developmental studies.

Bangs and whistles

Some phenomena do not fit neatly into either the language or the music category; they seem to have a foot in both camps. Consider the ‘talking drums’ of west and central Africa. Drummers communicate linguistic messages by mimicking the tones and syllabic rhythms of utterances in African tone languages. In these, the pitch pattern of a word is as much a part of its identity as its vowels and consonants. Changing a word’s pitch can entirely change its meaning — from ‘wing’ to ‘bag’, say.

The Yoruba people of Nigeria play the hourglass-shaped *dundun* drum. The leader uses the drum to ‘talk’ during musical performances. The Lokele of the upper Congo in central Africa use hollowed-out logs as large slit drums to communicate linguistic messages across wide stretches of jungle, as they can be heard far beyond the range of human voices. Drummed messages embedded in musical performance are understood by listeners familiar with the drum language, but can go completely unnoticed as language by a naive listener. Most important, the messages are not confined to a stock set of utterances. They can convey novel phrases — albeit not as efficiently as ordinary spoken language because ambiguities are created



D. PARKINS

by many words having the same tonal pattern. This problem is dealt with by placing words in longer, stereotyped poetic phrases.

Another speech surrogate is whistled language. Based on tone languages, oral whistles are used to convey the rhythmic and tonal pattern of syllables. Rhythm is cued by variations in the loudness of the whistle, and tonal patterns by the pitch. Whistled languages occur in

Africa, Asia and Central America. For example, the whistled language of the Hmong people of southeast Asia is based on their spoken language

and encodes the seven different tones that they use to distinguish word meaning in speech. Native listeners find it easy to understand despite minimal cues to vowel and consonant identity. Again, whistled speech can convey original linguistic utterances even though to the uninitiated the sound patterns may seem more like music. Ordinary languages or click languages are recognizable as speech, even when unintelligible; not so whistled languages.

The cognitive processes that enable the production and comprehension of talking drums and whistled speech are not well understood. They are almost completely unstudied by modern neuroscience. Yet they probably hold

important clues to how the biological building-blocks of language and music can be fluidly and dynamically reconfigured, rather than being exclusively bound to one domain or the other from birth.

Studies of non-Western music suggest that music is not an island in the brain. Intimations of deep links between music and language extend back to Plato, Charles Darwin and Ludwig Wittgenstein. Modern cognitive science is replacing speculation with research, and finding numerous links that bind these domains together as cognitive systems. These findings bear on a wide range of debates, from the ‘modularity’ of linguistic mechanisms to the evolutionary origins of music. We are a musical species as much as we are a linguistic one. By looking at cognition through both of these lenses, we may see deeper into the mechanisms that give our species its remarkable power to make sense of sound. ■

Aniruddh D. Patel is Esther J. Burnham senior fellow in theoretical neurobiology at the Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, California 92121, USA. He is author of *Music, Language, and the Brain* (Oxford University Press — see *Nature* **452**, 695–696; 2008 for a review).

For further reading see <http://tinyurl.com/6pscuj>.

“Studies of non-Western music suggest that music is not an island in the brain.”

GENOMICS

Protein fossils live on as RNA

Rajkumar Sasidharan and Mark Gerstein

Pseudogenes constitute many of the non-coding DNA sequences that make up large parts of genomes. Once considered merely protein fossils, it now emerges that some of them have active regulatory roles.

A central challenge in genome annotation is determining the function of sequences that do not encode proteins, but make up the overwhelming bulk of large genomes — some 99% in humans. A significant fraction of these sequences are pseudogenes, or fossils of ancient proteins, and although many of them are transcribed into RNA, they have hitherto been deemed 'junk'. However, given the abundance of pseudogenes, it is unlikely that they are useless. One function suggested for them is gene regulation, and RNA interference (RNAi) has been proposed as the mechanism for carrying this out. Six papers^{1–6}, including three in this issue (pages 793, 798 and 803), significantly expand the known scope of RNAi by describing the discovery of natural small interfering RNA (siRNA) sequences in mice and fruitflies, some of which are potentially transcribed from pseudogenes.

The textbook definition of a pseudogene is an inheritable genetic element that is similar to a functioning gene, yet is non-functional. But what is meant by non-functional is debatable — not transcribed, not translated, or not under control of a promoter sequence? Pseudogenes are similar to protein-coding genes because they are usually copied from a parent gene, either through unsuccessful duplication or by retrotransposition (whereby a gene is transcribed into RNA, which is then 'reverse-transcribed' back into DNA and inserted somewhere different in the genome). Because all this copying does not yield a normal, functioning protein, pseudogenes are usually identified by obvious 'disabilities' in their sequence, such as frameshifts or premature stops. They have been of interest because they provide records of ancient molecules encoded by the genome.

Although pseudogenes have generally been considered as evolutionary 'dead-ends', one of the surprises of genome sequencing has been how abundant they are: tens of thousands of pseudogenes are found in mammalian genomes (roughly the same number as protein-coding genes in all mammals sequenced so far)⁷. In addition, a large proportion of these sequences seem to be under some form of purifying selection⁸ — whereby natural selection

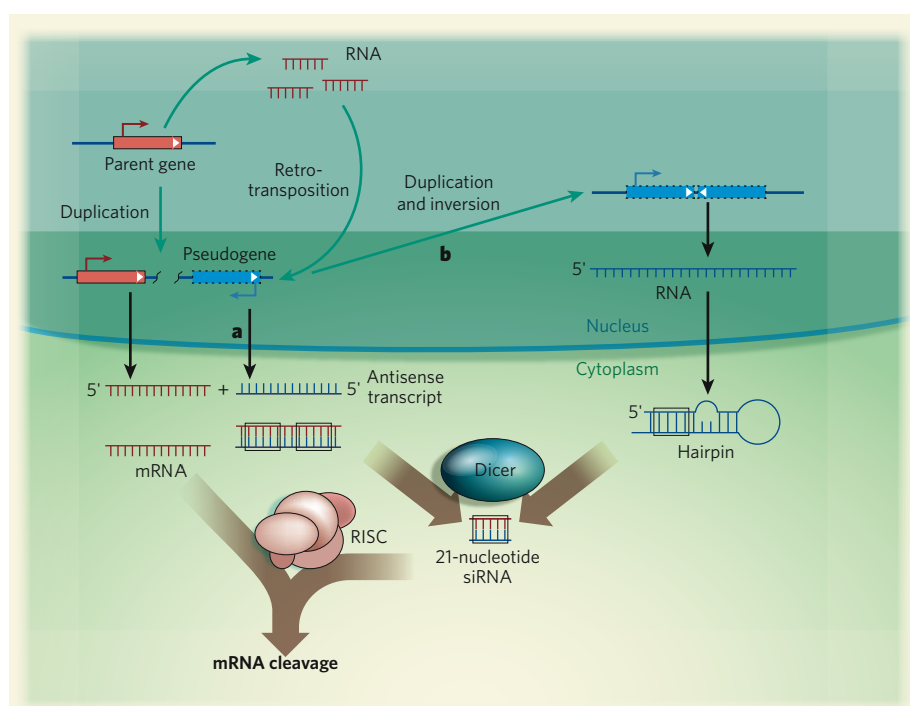


Figure 1 | Pseudogene-mediated production of endogenous small interfering RNAs (endo-siRNAs). Pseudogenes can arise through the copying of a parent gene (by duplication or by retrotransposition). **a**, An antisense transcript of the pseudogene and an mRNA transcript of its parent gene can then form a double-stranded RNA. **b**, Pseudogenic endo-siRNAs can also arise through copying of the parent gene as in **a** and then nearby duplication and inversion of this copy. The subsequent transcription of both copies results in a long RNA, which folds into a hairpin, as one half of it is complementary to its other half. In both **a** and **b**, the double-stranded RNA is cut by Dicer into 21-nucleotide endo-siRNAs, which are guided by the RISC complex to interact with, and degrade, the parent gene's remaining mRNA transcripts. The mRNA from genes is in red and that from pseudogenes is in blue. Green arrows indicate DNA rearrangements.

eliminates deleterious mutations from the population — and genetic elements under selection have some use. Finally, several large-scale genomic studies probing non-gene parts of the genome for biochemical activity have found many pseudogenes being transcribed and regulatory factors binding upstream of them. One such investigation, the ENCODE pilot project⁹, which looked at a representative 1% of the sequence of the human genome, found strong evidence for at least one-fifth of pseudogenes being actively transcribed.

These observations indicate that pseudogenes might not be purely dead relics of

past genes but could be resurrected for new biochemical activities. Indeed, functioning pseudogenes have been reported previously. For instance, in snails, a pseudogene is involved in translational control of the gene that codes for nitric oxide synthase¹⁰. And transcripts of the mouse pseudogene *makorin1-p1* have been proposed to inhibit degradation of their parent gene's mRNA, effectively enhancing its expression¹¹, although this observation has been debated. Nevertheless, a clear mechanism for the functioning of pseudogenes has been lacking. The six studies — four in flies^{1–4} and two in mice^{5,6} — provide such a direct pathway,

showing that pseudogene transcripts can act as natural siRNAs.

Broadly speaking, RNAi involves various types of small 'guide' RNA sequence regulating protein levels by targeting mRNA for degradation. Pseudogenic siRNAs provide two of the four categories posited by the six studies to organize the natural, or 'endo', siRNAs (Box 1).

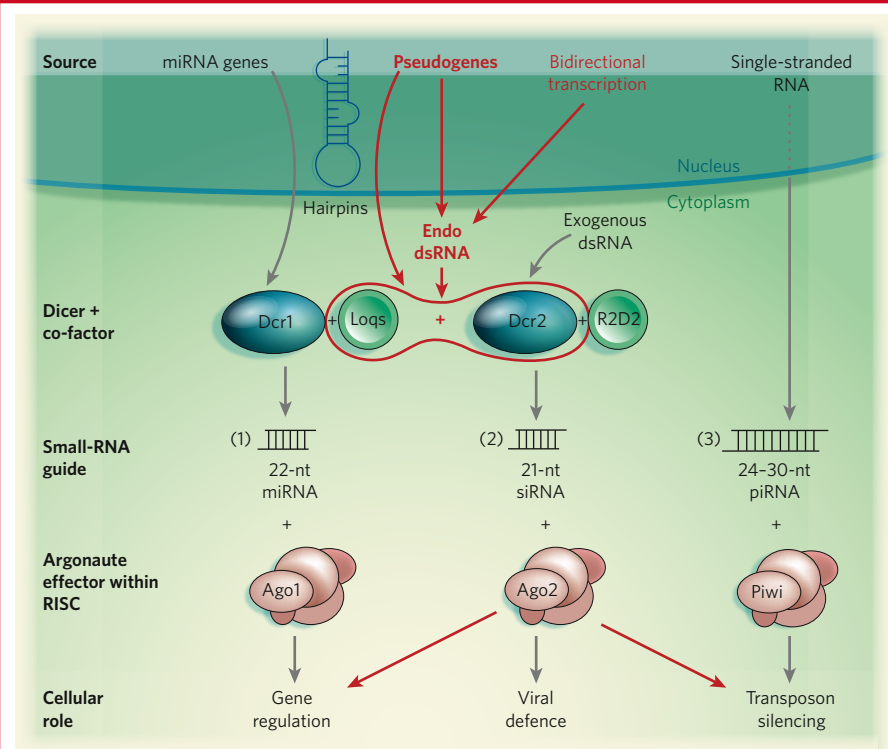
Endo-siRNAs in the first category mediate transposon silencing, which is typically a feature of Piwi-interacting RNAs (piRNAs). The studies were therefore careful to distinguish between endo-siRNAs associated with transposons and piRNAs on the basis of size (21–22 nucleotides versus 24–30) and Argonaute effector-protein partner (Ago2 versus Piwi). The second category of endo-siRNAs arise from bidirectional transcription of partially overlapping loci on opposite DNA strands^{1,12}. Studies in mice^{5,6} identify a few examples of these, and around 1,000 have been reported in flies¹, with their target genes consisting mainly of those with nucleic-acid functions, such as nuclease activity and transcription-factor binding¹².

The third category of siRNAs, which have been identified only in mice^{5,6}, are products of the interaction between a spliced mRNA transcript from a protein-coding parent gene and an antisense transcript from its pseudogene, which can be located far away from its parent gene, on the same or a different chromosome (Fig. 1a). Endo-siRNAs of the fourth category are closely related to those in the third. They arise from hairpin-shaped sequences, which in mice^{5,6} can come from inverted-repeat structures of pseudogenes (Fig. 1b). Here, the pseudogene also regulates its parent gene, but the double-stranded RNA precursor of the endo-siRNA comes from transcription of an inverted-repeat sequence, producing a hairpin. The reports show that mouse proteins affected by the third and fourth categories of endo-siRNAs are disproportionately involved in particular functions — such as regulating cytoskeletal dynamics — which indicates that their underlying pseudogene-mediated regulation has been explicitly selected for and is not simply caused by random pairing of transcribed genes and pseudogenes.

Hairpin precursors of endo-siRNAs have also been found in flies, but the evidence links them only weakly with inverted repeats of pseudogenes. Thus, most of the new data for pseudogenic siRNAs come from mouse rather than fly studies. One possible reason for this is that the mouse genome contains many more pseudogenes than the fly genome¹³. In fact, even compared with other metazoan organisms such as worms, flies are particularly poor in pseudogenes, possibly owing to pronounced genomic deletion processes known to occur in this organism¹⁴.

The scarcity of pseudogenes in flies makes their detection particularly difficult. Nevertheless, there is suggestive evidence for fly pseudogenes functioning as endo-siRNAs.

Box 1 | Small but significant



There are three main classes of small RNA, which generally differ in biogenesis, sorting and function¹⁹.

(1) MicroRNAs (miRNAs) mainly regulate genes involved in developmental processes. Specific miRNA genes encode mRNA-like primary transcripts that form hairpin structures, which are in turn excised by the enzyme Drosha (not shown) to form precursor miRNAs. In flies, further cleavage of these sequences by the Dicer enzyme Dcr1 and its specific co-factor Loqs yields mature miRNAs of ~22 nucleotides (nt). To carry out their function, miRNAs are incorporated into the RISC protein complex, which contains the effector protein

Ago1, a member of the Argonaute family.

(2) Conventional small-interfering RNAs (siRNAs) of ~21 nucleotides are produced through cleavage of double-stranded RNA (dsRNA) — in flies, by the Dicer enzyme Dcr2 and its co-factor R2D2 (refs 17, 19). These small RNAs bind to the Argonaute-family effector Ago2 and function in defence against external nucleic acids, such as synthetic dsRNAs or intermediates of viral replication.

(3) Discrete genomic loci give rise to single-stranded RNA sequences (ssRNA), which are then processed to ~27 nt Piwi-interacting RNAs (piRNAs). piRNA biosynthesis remains

somewhat ambiguous, but is known not to require Dicer. piRNAs bind to Piwi, another member of the Argonaute family that seems to be expressed only in germline cells. It is believed that these small RNAs function as master controllers of mobile genetic sequences called transposable elements²⁰.

In the figure, grey lines indicate known relationships, whereas red lines indicate new ones reported in the six papers¹⁻⁶. Clearly, the boundaries between the three small-RNA classes have been somewhat blurred by these reports. For details of how endo-siRNAs arise from pseudogenes, see Figure 1. **R.S. & M.G.**

First, an appreciable number (~30) have an inverted-repeat structure, associated with the formation of hairpins. Second, many of the sequences obtained by ultra-high-throughput sequencing of small RNAs in the fly coincide with DNA regions containing pseudogenes. In particular, a small but significant number of the 'reads' found using the Solexa sequencing technology^{1,4} can be intersected with some 70 pseudogenes, for an average of roughly 12 reads each. Finally, there is strong evidence that for several genes — particularly the β -esterase gene and its pseudogene — a duplicated pseudogene forms a functional complex with its parent gene, with regulatory consequences¹⁵.

Of course, to demonstrate the activity of pseudogenes conclusively, further experiments are needed. Deleting a pseudogene and demonstrating an effect on its potentially regulated parent gene would be most definitive. Also of great value would be studying the expression patterns of a potential endo-siRNA-producing pseudogene and its regulated parent gene across various tissues — data which should be generated by the ENCODE and modENCODE projects.

In addition to connecting RNAi with pseudogenes, the new studies¹⁻⁶ also blur the distinctions between the three 'traditional' classes of small RNA — siRNAs, piRNAs and

microRNAs (miRNAs) — which are distinct in their biogenesis and cellular roles (Box 1). The studies^{1–6} find that endo-siRNAs regulate transposons as piRNAs do; that, like miRNAs, they can arise from hairpins; and that, in flies, their processing involves a similar co-factor to the processing of miRNAs (Box 1).

This blurring of boundaries among different types of small RNA, together with the newly established links between siRNAs and pseudogenes, has interesting evolutionary implications. In plants, inverted duplications containing a protein-coding gene have been proposed¹⁶ as a mechanism to create new miRNAs. Thus, one can imagine a gene being copied (either by duplication or retrotranscription) and this copy then being duplicated (again) in inverted fashion. Given the ubiquitous nature of genomic transcription, the copy and its inverted duplicate could potentially be transcribed to a hairpin precursor of endo-siRNAs to regulate the parent gene.

As the function of the hairpin no longer has anything to do with encoding protein, its sequence, still under selection, can acquire frameshifts and stop codons, making it seem pseudogenic. One could even imagine its sequence drifting further and becoming gradually transformed into a miRNA gene, the sequence of which is much less similar to the gene encoding its target mRNA. So pseudogenes encoding endo-siRNAs might provide a crucial intermediate link to understanding the evolution of miRNA-mediated regulation¹⁷. Although speculative, the plausibility of this theory is bolstered by a recent survey¹⁸ of the genomic context of more than 300 human miRNA loci, which identified two that lie within pseudogenes.

Rajkumar Sasidharan and Mark Gerstein are in the Departments of Molecular Biophysics and Biochemistry, and Computer Science, Yale University, New Haven, Connecticut 06520, USA. e-mail: mark.gerstein@yale.edu

1. Czech, B. *et al.* *Nature* **453**, 798–802 (2008).
2. Ghildiyal, M. *et al.* *Science* **320**, 1077–1081 (2008).
3. Kawamura, Y. *et al.* *Nature* **453**, 793–797 (2008).
4. Okamura, K. *et al.* *Nature* **453**, 803–806 (2008).
5. Tam, O. H. *et al.* *Nature* **453**, 534–538 (2008).
6. Watanabe, T. *et al.* *Nature* **453**, 539–543 (2008).
7. Zhang, Z., Carriero, N. & Gerstein, M. *Trends Genet.* **20**, 62–67 (2004).
8. Zheng, D. *et al.* *Genome Res.* **17**, 839–851 (2007).
9. The ENCODE Project Consortium *Nature* **447**, 799–816 (2007).
10. Korneev, S. A., Park, J.-H. & O'Shea, M. J. *Neurosci.* **19**, 7711–7720 (1999).
11. Hirotsune, S. *et al.* *Nature* **423**, 91–96 (2003).
12. Okamura, K., Balla, S., Martin, R., Liu, N. & Lai, E. C. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1438 (2008).
13. Harrison, P. M., Milburn, D., Zhang, Z., Bertone, P. & Gerstein, M. *Nucleic Acids Res.* **31**, 1033–1037 (2003).
14. Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. *Nature* **384**, 346–349 (1996).
15. Balakirev, E. S., Anisimova, M. & Ayala, F. J. *J. Mol. Evol.* **62**, 496–510 (2006).
16. Allen, E. *et al.* *Nature Genet.* **36**, 1282–1290 (2004).
17. Chapman, E. J. & Carrington, J. C. *Nature Rev. Genet.* **8**, 884–896 (2007).
18. Devor, E. J. *J. Hered.* **97**, 186–190 (2006).
19. Matranga, C. & Zamore, P. D. *Curr. Biol.* **17**, R789–R793 (2007).
20. Brennecke, J. *et al.* *Cell* **128**, 1089–1103 (2007).

ATTOSECOND PHYSICS

An easier route to high harmony

Mark I. Stockman

The generation of ultrashort light pulses by atomic ionization and recombination doesn't come cheap. But by niftily exploiting the play of light on a nanostructured surface, it can be done on a table-top.

Extreme ultraviolet (EUV) radiation has great potential to be extreme not just in name, but in usefulness. It is the band of ultraviolet light with the shortest wavelength — around 5–50 nanometres, between 100 and 10 times shorter than that of visible light. In applications such as microscopy and lithography it can thus be used to probe and etch at tiny scales. What's more, this wavelength regime is that of many atomic resonances, making EUV light ideally suited for spectroscopic applications. On page 757 of this issue¹, Kim *et al.* detail a deft new way to produce EUV radiation — one that could be considerably more economical than previous approaches.

The way in which EUV radiation is currently generated is extremely fiddly. It starts with the amplification of light pulses from an oscillator, a source of laser light. These are used to drive the repeated ionization of noble-gas atoms. The electrons freed during this process are accelerated in the light field and, because the sign of the field reverses after half a cycle, re-collide with their parent atoms^{2,3}, releasing the electrons' surplus energy as light. The result is a sequence of ultrashort (attosecond) pulses that are themselves useful tools for high-time-resolution metrology^{4,5}. More detailed consideration of the process reveals that the spectrum of these pulses consists of a comb of 'high harmonics' — spectral lines at wavelengths equal to the wavelength of the driving field divided by some integer. (Owing to symmetries of the particular situation, only light corresponding to odd-integer divisors is generated in this case.) The highest-harmonic (shortest-wavelength) component of this spectrum can be selected by filtering to produce a single attosecond pulse at an EUV wavelength.

Things would be much simpler if EUV radiation could be produced directly from an oscillator — an ultrashort pulsed laser of relatively low intensity — without the need for sophisticated, complex and expensive amplifiers to produce a high-intensity optical field. Kim *et al.*¹ provide a distinct glimmer of an indication that such an approach could be viable. They illuminate an intricate, nanoscale gold antenna structure with light from a standard titanium–sapphire laser, with a wavelength of 800 nm. The interaction of the light with the antennas produces high harmonics right up to the 17th harmonic — whose wavelength of 47 nm lies within the EUV range. The optical intensity required to generate this light is, at 10^{11} W cm⁻², about 100 times less than in the traditional approaches.

The secret of the authors' success is the nanoscale behaviour of 'quasiparticles' known as surface plasmons. These packets of optical energy represent rapid oscillations of electron density that spring up in the surface regions of metal nanoparticles when bathed in an incident light field. If this incident light is of the right frequency, the surface plasmons can enter resonance, greatly increasing the local field intensity over that of the excitation wave. This phenomenon has a central role in, for example, surface-enhanced Raman scattering⁶, a spectroscopy and imaging technique that is sensitive enough to detect the presence of individual molecules adsorbed on a metal surface.

The extent of this field enhancement is determined by the nanoparticles' plasmonic resonance properties, which in turn depend mostly on the resistivity of the metal at the frequency of the optical light. Additional magnitude comes from geometric effects^{7–9}, both in narrow gaps between particles where there is a significant localization of optical energy, leading to the formation of 'gap plasmons', and also similarly around sharp tips, a phenomenon known as the lightning-rod effect. Nanoparticles have been specifically engineered in

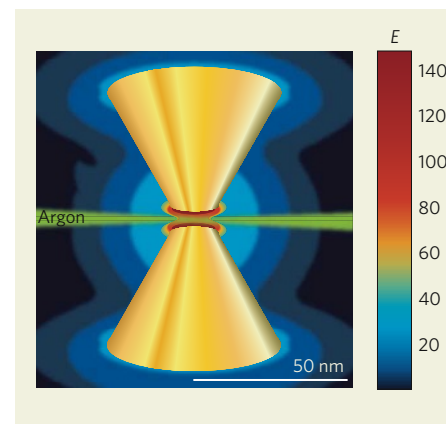


Figure 1 | Stripping on the table-top. The bow-tie-shaped gold nanoantennas used by Kim *et al.*¹ develop electric-field strengths in the gap through interactions with quasiparticles known as surface plasmons. (The field strength *E* is colour-coded; note that intensity is proportional to the square of the field strength.) When a beam of argon atoms (green) is directed towards the gap, the field strips them of electrons, which subsequently recombine — a process that results in the generation of high harmonics of the original light, including the sought-after extreme ultraviolet radiation. (Data and calculations courtesy of J. Aizpurua.)

quasi-two-dimensional, Mott–Hubbard charge transfer antiferromagnetic insulators in their undoped ground state. These last three properties are believed to be key to high-temperature superconductivity, and are about the only criteria on which you can find (almost) universal agreement among those trying to choose between the bespoke fashions hanging in the high- T_c theoretical closet.

However, observe in Figure 1 that the Fe ions, although nominally Fe^{2+} , analogous to Cu^{2+} , are tetrahedrally coordinated relative to the pnictide anions, as opposed to the square-planar symmetry of the copper oxide compounds. In the first-row transition metals — scandium to zinc — there are ten d -electron states (five described by orbital momentum), each of which can hold two electrons with one spin up and another down. We can play with these states to build various cationic configurations. A simple yardstick, called Hund's rule, helps build possible combinations in isolated atoms and ions. It says we have to start filling from the bottom, first occupying each orbital with an up-spin and then starting over again with spin-down, until all available d -electrons are consumed. Thus Fe^{2+} , with six electrons at large, will result in a ground state one electron in excess of a half-filled Hund's occupation distribution, and Cu^{2+} , with nine electrons to spend, will yield one electron fewer (a 'hole' or effective positive charge) than a filled d -orbital shell. Therefore, in a very crude sense, the new FeAs superconductors can be thought of as the electron analogues of the hole-transporting CuO complexes, and both measurements and theoretical studies bear this out.

The real situation is far more complex than just stated, and simple Hund's rule arguments are confounded by symmetry, position and overlap of neighbouring anions (O^{2-} , $\text{As}^{-(3-x)}$), and by Coulomb repulsion that tends to separate spins in otherwise 'Pauli-allowed' cation states from the next nearest cation neighbour. The trade jargon for these effects is 'crystal field splitting', 'hybridization' and 'Hubbard U', respectively. You can be assured each of these is currently undergoing intense exploration. All this notwithstanding, the simple Hund's rule picture that the ferrous pnictides and copper oxides are electron–hole 'duals' may not be simply fortuitous. It may be the reason that, after years of intense searching, nickel and cobalt complexes have not yielded high-temperature superconductors (at least not yet).

In fact, with T_c now at 55 K, are these ferrous pnictides truly 'high-temperature superconductors'? Simply answered, we don't know at present. But it is useful to remember that the expression 'high-temperature superconductivity' did not originate with Bednorz and Müller's paper¹ of 1986. Those who, like myself, are of mature years will recall that this description was coined as a result of studies^{7,8} in the 1960s that superconductivity mediated by electron–

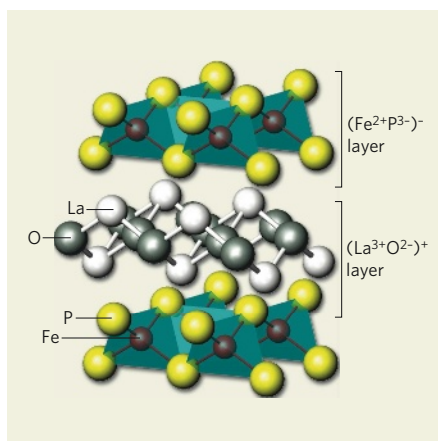


Figure 1 | The unit cell of LaOFeP. In this generic example³ of the family of lanthanum-series oxyfluoride ferrous pnictides, the overall cell charge is neutral but the individual layers are not, implying electron doping of the FeP layer. Note also that the P coordination of Fe is tetrahedral, not square planar as is the case for the high- T_c copper oxide perovskites. (Reproduced from ref. 3.)

phonon pairing would top out at around 30 K, and compounds showing anything above this value would be referred to as 'high-temperature materials'. Although the mechanism of high T_c in the copper oxide perovskites remains in question, we do have evidence⁹ in MgB_2 that electron–phonon coupling can achieve a transition temperature of 40 K. Is 55 K really that much higher?

Although most of the theoretical tailoring for the various ferrous pnictides is styled after fashions for the cuprate superconductors, one experimental study harks back to much earlier designs. Chen *et al.*¹⁰ report direct measurement of the superconducting energy gap and its temperature dependence in polycrystalline samples of $\text{Sm}(\text{O}_{0.85}\text{F}_{0.15})\text{FeAs}$, with $T_c = 42$ K. The technique used is called Andreev spectroscopy. This is a variant of tunnelling spectroscopy whereby, at a contact between a normal metal and a superconductor, an electron from the metal injected into the superconductor at energies lower than the superconducting gap gives rise to a superconducting pair (Cooper pair of electrons of opposite spin), which are subsequently spin-charge compensated by a 'reflection' of positive polarity (a hole) back into the normal metal. The resulting current–voltage dependence is a direct measurement of the superconducting pairing energy.

Astonishingly, Chen *et al.*¹⁰ find that their results best fit the time-honoured Bardeen–Cooper–Schrieffer (BCS) theory¹¹, the breakthrough in the mid-twentieth century that solved the riddle of superconductivity in all materials available up to that time. Although originally formulated to accommodate the pairing of electrons mediated by lattice vibrations (phonons), in its broadest sense the BCS framework can encompass pairing

of fermions in a boson field — perhaps even the 'flavours' found in neutron stars, quarks and gluons, giving rise to 'colour' superconductivity at the relatively low cosmological temperature of 10^9 (the units don't matter). So Chen and colleagues' identification of classic BCS behaviour does not rule out the possibility that some more exotic bosonic glue than phonons might be behind superconductivity in these ferrous pnictides.

Whenever a new superconductor with a T_c higher than 30 K appears on the scene, I inevitably get asked if it will bring applications closer. The question is perhaps more pertinent when the material involves particularly noxious elements such as arsenic. My answer is always "Just follow the money." If the pot at the end of the rainbow has enough gold inside (and so far it does not for applied superconductivity), the environmental issues can be overcome. I give you semiconductor manufacture and processing, which uses some of the most toxic compounds (including arsenides) in creation, yet is tolerated and brought under control because its financial return is in the trillions. Again, the units don't matter.

Will T_c in the pnictides continue to go up, and perhaps double or triple as happened in 1987–88? I doubt it. We've now been on standby for several months, and to my mind the best hope is that the discovery of pnictide high-temperature superconductivity will help us understand better the physics of the cuprates. The iron age has yet to dawn. ■

Paul M. Grant is at W2AGZ Technologies, 1147 Mockingbird Hill Lane, San Jose, California 95120, USA.

e-mail: w2agz@pacbell.net

1. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
2. Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908–910 (1987).
3. Kamihara, Y. *et al. J. Am. Chem. Soc.* **128**, 10012–10013 (2006).
4. Kamihara, Y., Watanabe, T., Hirano, M. & Hosono, H. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
5. Takahashi, H. *et al. Nature* **453**, 376–378 (2008).
6. Yang, J. *et al. Supercond. Sci. Technol.* **21**, doi:10.1088/0953-2048/21/8/082001 (2008).
7. McMillan, W. L. *Phys. Rev.* **167**, 331–344 (1968).
8. Allen, P. B. & Dynes, R. C. *Phys. Rev. B* **12**, 905–922 (1975).
9. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. *Nature* **410**, 63–64 (2001).
10. Chen, T. Y., Tesanovic, Z., Liu, R. H., Chen, X. H. & Chien, C. L. *Nature* doi:10.1038/nature07081 (2008).
11. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175–1204 (1957).

Correction

The News & Views article "Genomics: Protein fossils live on as RNA", by Rajkumar Sasidharan and Mark Gerstein (*Nature* **453**, 729–731; 2008), contains the following incorrect statement: "...reads' found using the Solexa sequencing technology^{1,4} can be intersected with some seven pseudogenes, for an average of roughly two reads each." In fact, these reads intersected with some 70 pseudogenes, for an average of roughly 12 reads each. Also, in the text of Box 1, 'nt' (nucleotide) was omitted from one passage, which should read "...to ~27 nt Piwi-interacting RNAs (piRNAs)." These corrections have already been made to the online versions of this article.

microRNAs (miRNAs) — which are distinct in their biogenesis and cellular roles (Box 1). The studies^{1–6} find that endo-siRNAs regulate transposons as piRNAs do; that, like miRNAs, they can arise from hairpins; and that, in flies, their processing involves a similar co-factor to the processing of miRNAs (Box 1).

This blurring of boundaries among different types of small RNA, together with the newly established links between siRNAs and pseudogenes, has interesting evolutionary implications. In plants, inverted duplications containing a protein-coding gene have been proposed¹⁶ as a mechanism to create new miRNAs. Thus, one can imagine a gene being copied (either by duplication or retrotranscription) and this copy then being duplicated (again) in inverted fashion. Given the ubiquitous nature of genomic transcription, the copy and its inverted duplicate could potentially be transcribed to a hairpin precursor of endo-siRNAs to regulate the parent gene.

As the function of the hairpin no longer has anything to do with encoding protein, its sequence, still under selection, can acquire frameshifts and stop codons, making it seem pseudogenic. One could even imagine its sequence drifting further and becoming gradually transformed into a miRNA gene, the sequence of which is much less similar to the gene encoding its target mRNA. So pseudogenes encoding endo-siRNAs might provide a crucial intermediate link to understanding the evolution of miRNA-mediated regulation¹⁷. Although speculative, the plausibility of this theory is bolstered by a recent survey¹⁸ of the genomic context of more than 300 human miRNA loci, which identified two that lie within pseudogenes.

Rajkumar Sasidharan and Mark Gerstein are in the Departments of Molecular Biophysics and Biochemistry, and Computer Science, Yale University, New Haven, Connecticut 06520, USA. e-mail: mark.gerstein@yale.edu

1. Czech, B. *et al.* *Nature* **453**, 798–802 (2008).
2. Ghildiyal, M. *et al.* *Science* **320**, 1077–1081 (2008).
3. Kawamura, Y. *et al.* *Nature* **453**, 793–797 (2008).
4. Okamura, K. *et al.* *Nature* **453**, 803–806 (2008).
5. Tam, O. H. *et al.* *Nature* **453**, 534–538 (2008).
6. Watanabe, T. *et al.* *Nature* **453**, 539–543 (2008).
7. Zhang, Z., Carriero, N. & Gerstein, M. *Trends Genet.* **20**, 62–67 (2004).
8. Zheng, D. *et al.* *Genome Res.* **17**, 839–851 (2007).
9. The ENCODE Project Consortium *Nature* **447**, 799–816 (2007).
10. Korneev, S. A., Park, J.-H. & O'Shea, M. J. *Neurosci.* **19**, 7711–7720 (1999).
11. Hirotsune, S. *et al.* *Nature* **423**, 91–96 (2003).
12. Okamura, K., Balla, S., Martin, R., Liu, N. & Lai, E. C. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1438 (2008).
13. Harrison, P. M., Milburn, D., Zhang, Z., Bertone, P. & Gerstein, M. *Nucleic Acids Res.* **31**, 1033–1037 (2003).
14. Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. *Nature* **384**, 346–349 (1996).
15. Balakirev, E. S., Anisimova, M. & Ayala, F. J. *J. Mol. Evol.* **62**, 496–510 (2006).
16. Allen, E. *et al.* *Nature Genet.* **36**, 1282–1290 (2004).
17. Chapman, E. J. & Carrington, J. C. *Nature Rev. Genet.* **8**, 884–896 (2007).
18. Devor, E. J. *J. Hered.* **97**, 186–190 (2006).
19. Matranga, C. & Zamore, P. D. *Curr. Biol.* **17**, R789–R793 (2007).
20. Brennecke, J. *et al.* *Cell* **128**, 1089–1103 (2007).

ATTOSECOND PHYSICS

An easier route to high harmony

Mark I. Stockman

The generation of ultrashort light pulses by atomic ionization and recombination doesn't come cheap. But by niftily exploiting the play of light on a nanostructured surface, it can be done on a table-top.

Extreme ultraviolet (EUV) radiation has great potential to be extreme not just in name, but in usefulness. It is the band of ultraviolet light with the shortest wavelength — around 5–50 nanometres, between 100 and 10 times shorter than that of visible light. In applications such as microscopy and lithography it can thus be used to probe and etch at tiny scales. What's more, this wavelength regime is that of many atomic resonances, making EUV light ideally suited for spectroscopic applications. On page 757 of this issue¹, Kim *et al.* detail a deft new way to produce EUV radiation — one that could be considerably more economical than previous approaches.

The way in which EUV radiation is currently generated is extremely fiddly. It starts with the amplification of light pulses from an oscillator, a source of laser light. These are used to drive the repeated ionization of noble-gas atoms. The electrons freed during this process are accelerated in the light field and, because the sign of the field reverses after half a cycle, re-collide with their parent atoms^{2,3}, releasing the electrons' surplus energy as light. The result is a sequence of ultrashort (attosecond) pulses that are themselves useful tools for high-time-resolution metrology^{4,5}. More detailed consideration of the process reveals that the spectrum of these pulses consists of a comb of 'high harmonics' — spectral lines at wavelengths equal to the wavelength of the driving field divided by some integer. (Owing to symmetries of the particular situation, only light corresponding to odd-integer divisors is generated in this case.) The highest-harmonic (shortest-wavelength) component of this spectrum can be selected by filtering to produce a single attosecond pulse at an EUV wavelength.

Things would be much simpler if EUV radiation could be produced directly from an oscillator — an ultrashort pulsed laser of relatively low intensity — without the need for sophisticated, complex and expensive amplifiers to produce a high-intensity optical field. Kim *et al.*¹ provide a distinct glimmer of an indication that such an approach could be viable. They illuminate an intricate, nanoscale gold antenna structure with light from a standard titanium-sapphire laser, with a wavelength of 800 nm. The interaction of the light with the antennas produces high harmonics right up to the 17th harmonic — whose wavelength of 47 nm lies within the EUV range. The optical intensity required to generate this light is, at 10^{11} W cm⁻², about 100 times less than in the traditional approaches.

The secret of the authors' success is the nanoscale behaviour of 'quasiparticles' known as surface plasmons. These packets of optical energy represent rapid oscillations of electron density that spring up in the surface regions of metal nanoparticles when bathed in an incident light field. If this incident light is of the right frequency, the surface plasmons can enter resonance, greatly increasing the local field intensity over that of the excitation wave. This phenomenon has a central role in, for example, surface-enhanced Raman scattering⁶, a spectroscopy and imaging technique that is sensitive enough to detect the presence of individual molecules adsorbed on a metal surface.

The extent of this field enhancement is determined by the nanoparticles' plasmonic resonance properties, which in turn depend mostly on the resistivity of the metal at the frequency of the optical light. Additional magnitude comes from geometric effects^{7–9}, both in narrow gaps between particles where there is a significant localization of optical energy, leading to the formation of 'gap plasmons', and also similarly around sharp tips, a phenomenon known as the lightning-rod effect. Nanoparticles have been specifically engineered in

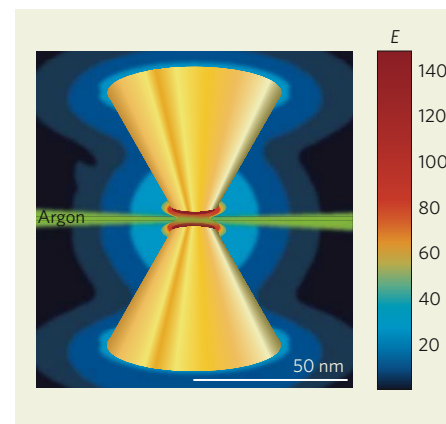


Figure 1 | Stripping on the table-top. The bow-tie-shaped gold nanoantennas used by Kim *et al.*¹ develop electric-field strengths in the gap through interactions with quasiparticles known as surface plasmons. (The field strength E is colour-coded; note that intensity is proportional to the square of the field strength.) When a beam of argon atoms (green) is directed towards the gap, the field strips them of electrons, which subsequently recombine — a process that results in the generation of high harmonics of the original light, including the sought-after extreme ultraviolet radiation. (Data and calculations courtesy of J. Aizpurua.)

configurations such as nanolenses¹⁰ and bow-tie nanoantennas¹¹ to take advantage of all three mechanisms: resonance, gap plasmons and the lightning-rod effect.

The context for Kim and colleagues' experiments¹ is an array of gold bow-tie nanoantennas on a sapphire substrate. In the nanoscale gaps of the bow-tie, the local optical intensity can be enhanced in an idealized case by up to 10,000 times (Fig. 1). In reality, the enhancement is limited by intrinsic optical losses in gold, imperfections in the structure's geometry, and the defects and polycrystalline nature of the gold's crystal lattice. But if the local field intensity is increased by just 100 times, it reaches 10^{13} W cm⁻², sufficient for relatively efficient generation of EUV radiation. Because this field is extremely locally concentrated in the gaps, it does not damage the sapphire substrate. Equally, the high dielectric permittivity of the gold nanoparticles means that the field cannot penetrate far into their surface.

The authors immerse their bow-tie assembly in a jet of argon gas. From this point, everything proceeds broadly analogously to the traditional production of EUV light. First, the high local field created by the plasmonic enhancement detaches an electron from the argon atoms as they pass through the bow-tie gaps. This electron is accelerated by the optical fields along its trajectory, and collides and recombines with the atom. The oscillating optical field causes the collisions to repeat periodically, releasing energy at high-harmonic frequencies of the original light.

There are two principal distinctions between this process and the generation of high harmonics by intense optical pulses. First, the intensity threshold is considerably lower, as touched on earlier. Second, the incident optical pulses are relatively tightly focused, and so the EUV light is generated in just one or a few nanogaps of the array and is not highly directed. (Directedness would stem from the interference of radiation from many bow-ties.)

This new method of short-wavelength light generation will open doors in imaging, lithography and spectroscopy on the nanoscale. It could expand the range of techniques that exploit local-field enhancement, such as near-field scanning optical microscopy, from the infrared and optical wavelengths to the EUV and, ultimately, as the plasmonic quality of the nanoantennas improves, X-ray range. This much shorter wavelength will benefit techniques such as core spectroscopy and nanoscale X-ray crystallography.

With less-tight focusing, it will also become possible to generate EUV radiation in many nanoantennas simultaneously. The resulting spatially coherent, laser-like light could have applications in many areas: in macroscopic spectroscopies, in screening for structural defects in materials and, extended to X-ray and even γ -ray wavelengths, in the detection of minute amounts of fissile materials in the arenas of public security and defence. ■

Mark I. Stockman is in the Department of Physics and Astronomy, Georgia State University, University Plaza, Atlanta, Georgia 30303-3083, USA.

e-mail: mstockman@mail.phy-astr.gsu.edu

1. Kim, S. *et al.* *Nature* **453**, 757–760 (2008).
2. Corkum, P. B. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
3. Chang, Z., Rundquist, A., Wang, H., Murnane, M. M. & Kapteyn, H. C. *Phys. Rev. Lett.* **79**, 2967–2970 (1997).
4. Paul, P. M. *et al.* *Science* **292**, 1689–1692 (2001).

5. Hentschel, M. *et al.* *Nature* **414**, 509–513 (2001).
6. Kneipp, K., Moskovits, M. & Kneipp, H. (eds) *Surface-Enhanced Raman Scattering: Physics and Applications* (Springer, Heidelberg, 2006).
7. Krenn, J. R. *et al.* *Phys. Rev. Lett.* **82**, 2590–2593 (1999).
8. Mühlischlegel, P., Eisler, H.-J., Martin, O. J. F., Hecht, B. & Pohl, D. W. *Science* **308**, 1607–1609 (2005).
9. Romero, I., Aizpurua, J., Bryant, G. W. & de Abajo, F. J. G. *Opt. Express* **14**, 9988–9999 (2006).
10. Li, K., Stockman, M. I. & Bergman, D. J. *Phys. Rev. Lett.* **91**, 227402 (2003).
11. Schuck, P. J., Fromm, D. P., Sundaramurthy, A., Kino, G. S. & Moerner, W. E. *Phys. Rev. Lett.* **94**, 017402 (2005).

DEVELOPMENTAL BIOLOGY

Order in the lung

David Warburton

Given the lung's thousands of branching airways, its development might be expected to be a highly complex process. Yet a surprisingly simple picture now emerges of when, where and in what order these branches form.

Elaborate branching is everywhere in nature. From riverbeds to oilfields, from trees to blood vessels, branching connects the large to the small. The lung is also a prime example of a reproducible branching system, allowing gas to be transported from the air to tissues deep within an animal. Without it — or without the simpler branched ducts found in less complex organisms — oxygen transport by diffusion probably would have limited the evolution of terrestrial animals to less than one millimetre in size. But how does such a sophisticated network develop? Metzger *et al.*¹ (page 745 of this issue) provide a remarkable, yet simple picture that explains the orderly development of the more than a million branches in the mammalian lung.

In mammals, air enters through the nasal and oral cavities and passes through the larynx and trachea before reaching the lung. The trachea branches into two primary bronchi, which, within the lung, further branch into secondary and tertiary bronchi and finally into bronchioles. To investigate the sequence of events leading to this complex, yet highly reproducible network of branches, Metzger *et al.* studied the early bronchial tree in three dimensions by examining chemically fixed lung tissue from mouse embryos using microscopy.

The authors parse bronchial branching beyond the primary branch into three geometrical modes, which they call domain branching, planar bifurcation and orthogonal bifurcation. In domain branching, daughter branches form in rows along the parent branch, like bristles on a bottle brush. This branching mode forms the main secondary branches. Next, planar bifurcation is used for the formation of tertiary and later-generation branches; this mode is characterized by the splitting of a branch tip into two. Finally, orthogonal bifurcation involves two rounds of branching. Both rounds involve

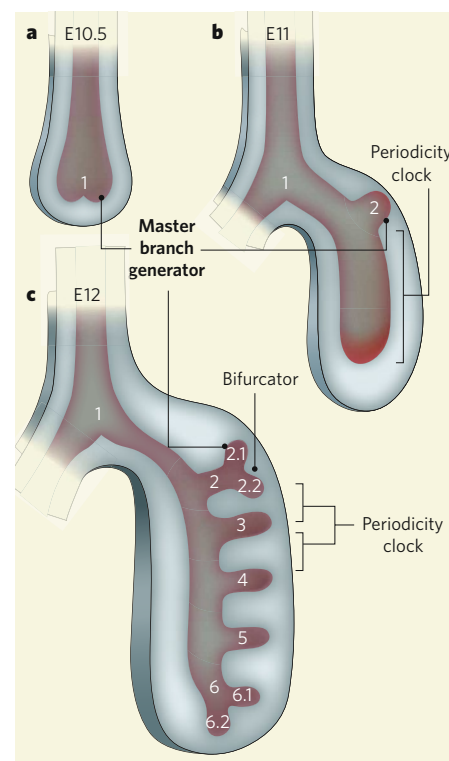


Figure 1 | A master and three slaves. Studying early lung morphogenesis in the mouse embryo, Metzger *et al.*¹ show that this organ's airways form in a sequential manner in three series of events, or subroutines, which are all driven by one master branch generator. **a**, At embryonic day (E) 10.5, the primary bronchial branch (1) forms, followed by **(b)** the development of the left upper-lobe branch (2) by E11. **c**, The first two segmental branches of the left upper-lobe branch (2.2 and 2.3) and the subsequent formation of branches 3–6 occur at E12. The master branch generator is active throughout these events, and the inferred sites of action of the periodicity clock and bifurcator subroutines are shown. At E12, the rotator subroutine has not yet begun to function.

Q&A

CONDENSED-MATTER PHYSICS

Optical lattices

Markus Greiner and Simon Fölling

Optical lattices have rapidly become a favoured tool of atomic and condensed-matter physicists. These crystals made of light can be used to trap atoms at very low temperatures, creating a workshop in which to pore over and tinker with fundamental properties of matter.

Why make an optical lattice?

Imagine you are trying to understand a complex quantum-physical phenomenon such as high-temperature superconductivity, just one of many intriguing effects that occur in the 'quantum gas' of electrons that pervades a solid crystal. The crystal lattice of a solid has tiny dimensions — atoms are spaced about a tenth of a nanometre apart — and it would be great to have an enlarged version to enable us to look at its physics more closely. An optical lattice affords just that possibility: it is a crystal formed by interfering laser beams, with a typical dimension about 1,000 times larger than that of a conventional crystal. Ultracold atoms in the lattice play the part of electrons in the solid; they tunnel quantum-mechanically between lattice sites just as single or paired electrons (Cooper pairs) tunnel through the periodic potential wells created by positive ions in crystalline materials (Fig. 1).

Why not just use a real crystal?

Real solid materials are incredibly complex. They have an involved band structure of allowed energy states, and the Coulomb interactions of electrons are difficult to account

for. Disorder and the inevitable effects of vibrations of the crystal lattice add to the intricacy. It is not desirable, and usually impossible, to take account of all these effects simultaneously. The theorists' approach is to construct highly simplified models that focus on particular aspects of the system. But often even these cannot be solved reliably. That is particularly true of models involving fermions — the group of particles to which electrons, protons, neutrons and many atoms belong. And so a divide has opened up: on the one hand, real materials cannot be sufficiently described by theorists; on the other, experimentalists cannot test the (sometimes contradictory) theoretical results obtained from simple models.

How can optical lattices help?

They provide a way of realizing the simplified models of condensed-matter theory in experimental practice. Optical lattices thus implement Richard Feynman's pioneering idea of 'quantum simulation' — using one quantum system to investigate another. Feynman introduced the concept because a correct numerical description of a quantum system requires resources that scale exponentially

with the number of particles involved — one of the problems confronting theorists today. Using results from optical lattices, theorists want to test which of their models they can best rely on to, say, construct a phase diagram of a condensed-matter system, or assess the evolution of a physical parameter. In addition, optical lattices provide ways to control various factors, such as the strength of interatomic interactions, band structure, spin composition and levels of disorder, more easily than in real crystals — often even dynamically during the course of a single experiment.

Are there limits to what you can simulate with an optical lattice?

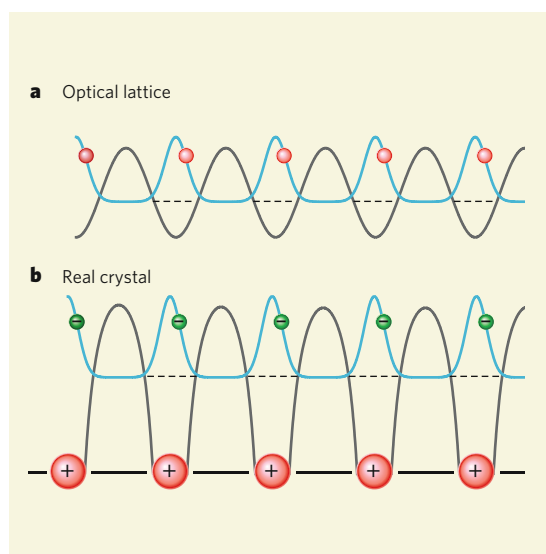
We can't realistically use them to simulate very complicated crystals, such as superconducting ceramics, in their entirety. But that's not the goal. To understand phenomena such as high-temperature superconductivity, we want to find the simplest possible system that shows the effect. Consequently, it is not a bad idea to start from simple models and add complexity step by step.

How exactly do you make an optical lattice?

Lattice potentials are created by making two coherent laser beams propagating in opposite directions interfere with each other. The result is a standing wave with a periodic pattern of dark and bright stripes. The light induces an electric dipole moment in the atoms of the ultracold gas, modifying their energy. Depending on the frequency of the light, atoms are pulled towards either the bright or the dark regions and are therefore confined to small areas in space. By using additional lasers from different directions, two- or even three-dimensional lattice structures can be constructed (Fig. 2).

Can you make a lattice for any kind of atom?

Pretty much, although alkali atoms from group I of the periodic table are generally easiest to work with — their single valence electron gives them

**Figure 1 | Crystal simulation.**

Ultracold atoms in an optical lattice can simulate condensed-matter phenomena that usually occur only in the 'electron gas' of a solid-state crystal. In an optical lattice (a), atoms are trapped in a sinusoidal potential well (grey) created by a standing-wave laser beam. The atoms' wavefunctions (blue) correspond to those of valence electrons in a real crystal (b). Here, the periodic potential is caused by the attractive electrostatic force between the electrons (–) and the ions (+) forming the crystal. The motion and interaction of the particles, whether ultracold atoms or electrons, determine the physics of the material. Thus, for example, superfluidity in a gas of ultracold atoms corresponds to superconductivity in an electron gas.

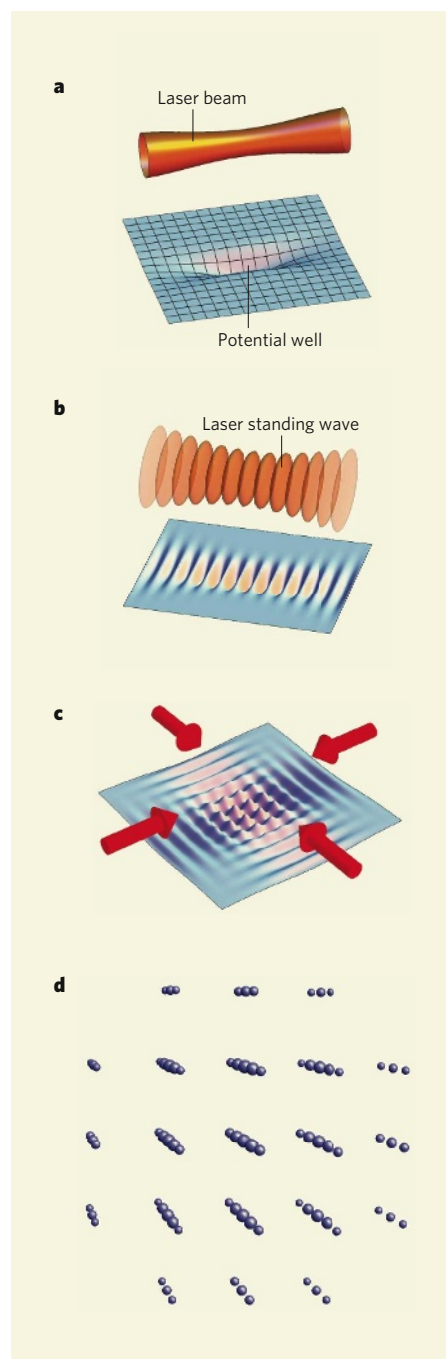


Figure 2 | Potential lattice. Laser light creates the peculiar potential landscapes of optical lattices. **a**, Laser light creates a repulsive or an attractive potential that is proportional to the laser's intensity along its axis of propagation. **b**, By allowing two counter-propagating laser beams to interfere, a sinusoidal standing wave can be formed. Ultracold atoms can be trapped in the potential minima that occur every half-wavelength, separated from the neighbouring minimum by a potential 'wall'; this is the basis of a one-dimensional optical lattice. **c**, **d**, Adding more laser beams at right angles to the first one creates a two-dimensional (**c**) and finally a three-dimensional cubic 'crystal' in which atoms (**d**) are trapped in the minima of the lattice potential. Real optical lattices can have millions of sites, and by changing the angles and wavelength of the laser beams, different lattice geometries can be created.

particularly suitable optical properties. At the very low energy scales (temperatures) at which optical-lattice models operate, most atoms are actually very similar. The most important property is whether they are bosons (with integer spin) or fermions (with half-integer spin). Many bosons can exist in the same quantum state, but quantum mechanics absolutely forbids this for fermions, according to the rule known as the Pauli exclusion principle. Whether a particular atomic isotope is bosonic or fermionic depends on the number of its (fermionic) constituents: protons, neutrons and electrons. If this number is even, the total spin is an integer, and the atom is a boson; if it is odd, the total spin is a half-integer, and the atom is a fermion. It's easier to work with bosons: one consequence of the Pauli exclusion principle is that fermions are tougher to cool. Today, optical-lattice experiments are being pursued with many different atoms — bosonic rubidium-87, sodium-23, potassium-39 and caesium-133; and fermionic potassium-40, lithium-6 and strontium-87.

How do you put the atoms into the lattice?

Usually, by approaching the problem from the other direction: you put the lattice into the atoms. First, you create a Bose–Einstein condensate — an ensemble of bosons all in the same quantum state — or a cold gas of fermionic atoms. Then you slowly ramp up the lasers to create the periodic lattice potential, and the atoms reorder to adapt to their new environment. Similarly, the whole 'crystal' can be removed from the atoms simply by ramping down the lasers, thus liberating the atoms into free space once more.

How did the work with optical lattices start?

The first generation of optical lattices came on the scene in the 1990s. They were mainly used for the laser-cooling of atoms, as the lattice potential increases the efficiency of some methods of optical cooling. Still, the temperatures at that point were too high and the occupation numbers too low (many lattice sites were left unfilled) for many-body quantum physics to feature. Nevertheless, fundamental phenomena such as Bloch oscillations of atoms within the lattice and the Bragg scattering of light on the atoms could already be observed.

When did people start using lattices to investigate many-body physics?

A major boost came with the creation of the first Bose–Einstein condensates in the mid-1990s. This meant the availability of a dense and extremely cold sample of atoms on which a lattice could simply be superimposed. It became possible to study superfluid behaviour in the lattice, emulating for example the physics of arrays of Josephson junctions — tunnel junctions in superconductors that have important applications in quantum-mechanical circuits. But experiments with complex many-body

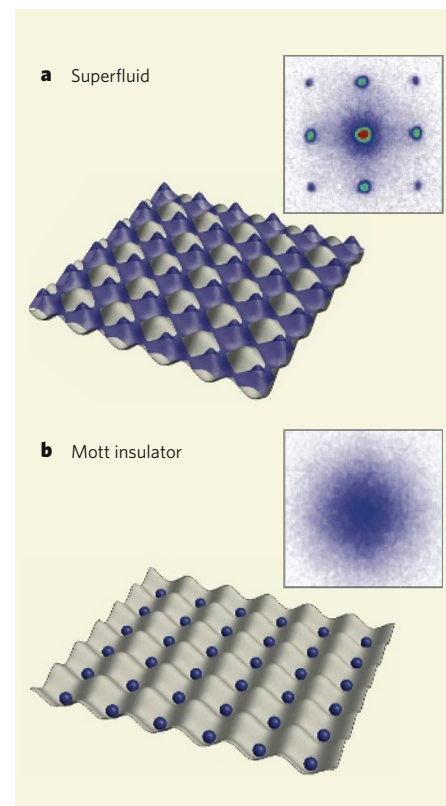


Figure 3 | Superfluids and insulators. In a quantum simulation of a bosonic Hubbard model in an optical lattice, two distinct ground states exist. **a**, For weak interaction strengths the ground state is a superfluid. The atoms have a common macroscopic wavefunction (purple wave), with particles fully delocalized throughout the available space. After the atoms are released from the lattice, the coherent wave nature becomes evident in the fact that a distinct diffraction pattern arises (inset). **b**, If the interactions become very strong, the ground state is a 'Mott insulator' in which each particle becomes localized to one specific lattice site. In that state, the number of atoms in each site is fixed, and phase coherence, as well as superfluidity, disappears.

states, relevant for simulating condensed-matter systems, started only around 2000, first with Bose–Einstein condensates in three-dimensional lattices, and later with ultracold Fermi gases.

What are three-dimensional lattices good for?

This geometry turns out to be a virtually perfect rendering of the Hubbard model, a basic model that describes particles in a crystalline lattice. It takes into account the atoms' mutual repulsion or attraction and the rate at which they hop from site to site. Experiments with bosonic atoms in optical lattices confirmed results arrived at using models, in particular that ultracold bosons undergo a transition between two very distinct quantum states. The first is the Bose–Einstein condensate, a superfluid state in which the gas can move without friction and particles are delocalized over

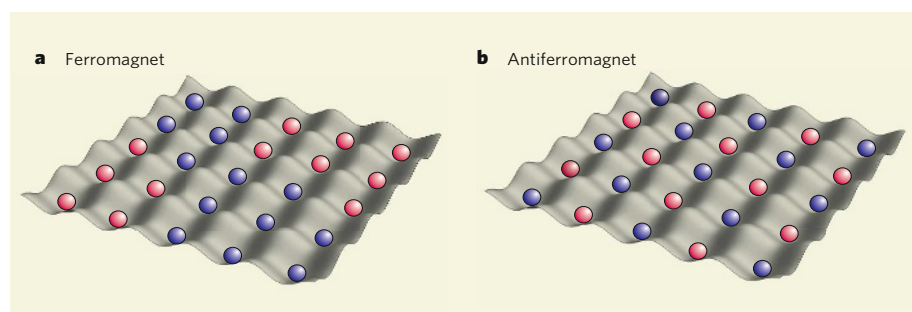


Figure 4 | Magnetic effects. When atoms in two different spin states (purple or red) are simultaneously present in the lattice, they can either favour or disfavour identical spins on neighbouring sites. **a**, In the first case, they distribute in such a way that compact areas of identical spins form — magnetic ‘domains’, which are the hallmark of ferromagnetism in solids. **b**, Conversely, in the antiferromagnetic state, a chessboard pattern of alternating spins develops. The underlying force driving the system into one of these states might be direct, long-range interactions, for example through the magnetic dipole of the atoms. Surprisingly, it can also arise from the tunnelling itself in combination with the peculiar quantum statistical properties of bosons and fermions. This ‘quantum magnetism’ does not require any direct magnetic interaction between the particles, and should be observable in an optical lattice implementing a simple fermionic Hubbard model.

the whole lattice. When the lattice ‘depth’ (the height of the potential barriers) is increased, the role of interactions between particles increases compared with that of the tunnelling. Eventually, the interactions bring the system into what is known as a Mott insulator state, in which free movement becomes impossible and each particle is localized to one lattice site only (Fig. 3).

Can optical lattices go beyond merely confirming theory?

So far, experimenters have mostly studied the Bose–Hubbard model: thanks to numerical simulations, we understand this model well, so it’s easy to compare experimental and numerical results. But there’s a lot of physics in boson interactions that isn’t covered by this ‘plain vanilla’ model, and that is not so well understood theoretically. For example, introducing more than one atomic spin state leads to magnetic effects. As different types of interactions compete, each favouring a different type of magnetic ordering, exciting new states such as ‘spin liquids’ are expected to arise. Then there are fermionic Hubbard models. As mentioned before, fermions behave fundamentally differently from bosons because of the Pauli exclusion principle. To form a superfluid, which requires all the particles to be in the same, coherent quantum state, fermions must ‘pretend’ to be bosons by, for example, teaming up to form Cooper pairs. Several laboratories are working on reproducing the behaviour of fermions in an optical lattice. Promising results, such as the creation of insulating and superfluid behaviours, are already being reported.

Why should we care about the fermionic Hubbard model?

One reason is what it might tell us about high-temperature superconductivity. This puzzling phenomenon — electrical conduction without resistance at temperatures of up to 130 kelvin — occurs in complex cuprate (copper oxide) materials, and has recently been reported in

a second, unrelated class of material. Despite 20 years of effort, there is still no consensus as to what causes either the formation of the ‘d-wave’ Cooper pairs that seem to lie behind the superconductivity, or other anomalous properties of these materials. The Hubbard model could be a good place to start: high-temperature superconducting phases seem always to be found near phases with antiferromagnetic order, in which particles of opposing spin states arrange themselves in a chessboard pattern (Fig. 4). This is a typical ground state of a fermionic Hubbard model.

So the Hubbard model holds the key to high-temperature superconductivity?

There are many reasons to believe that the Hubbard model contains most — but not all — of the ingredients necessary for understanding high-temperature superconductivity in the cuprates. A full explanation is likely to require additional phenomena, such as the interactions between electrons and lattice vibrations (phonons), and inhomogeneities at the mesoscopic level, although there is no consensus about which, if any, of these ingredients are important. But that is just where the remarkable control over ultracold atoms in an optical lattice comes into its own — we expect that we can approach the problem by realizing the bare Hubbard model first and then adding the other ingredients in a controlled way. This really would be a quantum simulator as Feynman first envisaged it.

How might we add to this quantum-simulation tool-box in the future?

A current focus is on extending the technique to further atomic species, each with different interaction properties, as well as on decreasing the temperature of the trapped atoms still further through new cooling techniques. An intense line of research is also to introduce long-range interactions into optical lattices by trapping, not atoms, but molecules with a dipolar

ground state that interact through electric dipole–dipole forces. And then there are efforts to improve how we can analyse the quantum state, for example by directly reading out every single lattice site or by probing characteristic quantum fluctuations of the whole ensemble.

Are there applications for optical lattices beyond condensed-matter simulations?

Because as few as one or two atoms can be isolated in each of the small traps, the traps can also act as miniature ‘test tubes’ in which atomic interactions can be studied with high precision. Even ‘micro-chemistry’ can be performed, with the ability, for example, to precisely control the formation of a molecule from two atoms. Equally, with just one atom in each site, interactions can be completely suppressed. That is seen as a way of improving the precision of optical atomic clocks still further. These create highly accurate time references through the extremely fast oscillations that occur between the energy states of the atoms. Those energy states are disturbed by atom–atom interactions that can be suppressed in an optical lattice. And last but not least, there are intriguing proposals to use atoms trapped at single lattice sites as ‘qubits’ for a so-called universal quantum computer.

How would a universal quantum computer work?

The first step would be to create large-scale entanglement between all atoms in a lattice, by letting each atom interact only with its next neighbour in a way that depends on its state. This technique has already been demonstrated experimentally. In principle, a universal quantum computer can be realized just by reading out and manipulating these single entangled qubits following such an entanglement step. The challenge is to do all the steps — entanglement, manipulation, readout and error correction — with high enough fidelity. We might have to wait years to see a useful universal quantum computer. But in the meantime, quantum simulators in the spirit of Feynman are already a specialized type of quantum computer: simulated condensed-matter models in optical lattices look set to become the first practically useful application of a quantum-computational system. ■

Markus Greiner and Simon Fölling are in the Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA.

e-mails: greiner@physics.harvard.edu; foelling@physics.harvard.edu

FURTHER READING

Feynman, R. P. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

Bloch, I. *Nature Phys.* **1**, 23–30 (2005).

Auerbach, A. *Interacting Electrons and Quantum Magnetism* (Springer, Berlin, 2006).

Bloch, I., Dalibard, J. & Zwerger, W. Preprint at <http://arxiv.org/pdf/0704.3011> (2007).

The determination of the structure of Saturn's F ring by nearby moonlets

Carl D. Murray¹, Kevin Beurle¹, Nicholas J. Cooper¹, Michael W. Evans¹, Gareth A. Williams¹ & Sébastien Charnoz²

Saturn's narrow F ring exhibits several unusual features that vary on timescales of hours to years. These include transient clumps, a central core surrounded by a multistranded structure and a regular series of longitudinal channels associated with Prometheus, one of the ring's two 'shepherding' satellites. Several smaller moonlets and clumps have been detected in the ring's immediate vicinity, and a population of embedded objects has been inferred. Here we report direct evidence of moonlets embedded in the ring's bright core, and show that most of the F ring's morphology results from the continual gravitational and collisional effects of small satellites, often combined with the perturbing effect of Prometheus. The F-ring region is perhaps the only location in the Solar System where large-scale collisional processes are occurring on an almost daily basis.

Saturn's F ring, which lies 3,400 km beyond the edge of the main ring system, has presented many unusual aspects and associated dynamical phenomena since its discovery by Pioneer 11 (ref. 1). Subsequent imaging by the Voyager spacecraft^{2,3} showed a narrow ring with multiple components and a twisted, braid-like appearance, containing clumps and radial discontinuities. Voyager radio science^{4,5} and photopolarimeter⁶ occultation data suggested a core ~ 1 km wide surrounded by an envelope 50 km wide. Surprisingly, despite the tendency for narrow rings to spread owing to collisions, and the fact that it should be significantly perturbed by its shepherding satellites, Prometheus and Pandora, the F ring is well approximated (residuals ~ 5 km) by a narrow, uniformly precessing ring of material⁷.

Early imaging by the Cassini spacecraft⁸ confirmed the multi-stranded nature of the F ring and displayed streamers and associated channels of ring material gravitationally affected by Prometheus as predicted by numerical models⁹. Additional observations and modelling showed¹⁰ how the Prometheus streamer-channel cycles and their shearing cause characteristic regular radial and longitudinal variations in the F ring every 14.7-h orbital period. However, the discovery that the apparent multiple strands of the F ring had the structure of a kinematic spiral¹¹ was the first evidence that additional objects were involved in producing some of the F ring's more unusual features.

Two possible explanations were suggested¹¹ for the formation of the spiral structures detected in mosaics of the F ring obtained in November 2004, April 2005 and May 2005. The first involved ring material being gravitationally scattered during the close approach of an object with the mass of Prometheus. However, no such object has been observed and Prometheus itself does not enter the ring's core. The second explanation proposed that particles were pushed out of the ring when a small object passed through it. Additional objects (S/2004 S 3, S/2004 S 4 and S/2004 S 6 with estimated diameters ~ 5 km) have been detected by Cassini^{8,12} in the vicinity of the F ring, and one of them, S/2004 S 6, was suggested as a possible impactor because of the correspondence between its F-ring crossing points and the location of the spirals¹¹. It had been independently suggested^{13,14} that collisions could generate some of the bright clumps in the F ring, but the impactors were thought to be ~ 10 -cm-sized interplanetary meteoroids. The detection of 13 opaque or nearly-opaque objects during stellar occultation experiments using the Cassini Ultraviolet Imaging Spectrograph

and Visual and Infrared Mapping Spectrometer¹⁵ suggests the existence of objects with diameters ranging from 27 m to 9 km lying within ~ 10 km of a core that is ~ 40 km wide; at least one of these objects is opaque enough to be a moonlet rather than a temporary clump of material. The Voyager charged-particle data had already provided evidence¹⁶ that a 2,000-km-wide moonlet belt surrounded the F ring, and this led to interpretations of the bright clumps as products of the disruptive collisions of loosely bound parent bodies^{17,18}.

All the evidence suggests that the F-ring region is a unique dynamical environment in which gravitational perturbations combine with physical collisions to alter the F ring's appearance dramatically on several timescales. However, the actual mechanisms and objects that change the ring have remained unknown. Here we report on the analysis of a series of images obtained by the Cassini Imaging Science Subsystem¹⁹ in 2006 and 2007 (see Supplementary Table 1). These show clear evidence for collisional events, the creation and evolution of associated 'jets' of material, the gravitational signatures of ~ 10 -km-diameter satellites embedded in the core, the perturbing effects of smaller objects, and direct observations of a narrow (radial width ~ 1 km) ring component and associated objects.

Colliding objects

Figure 1 illustrates how the morphology of the F ring changed between 2006 December 23 and 2007 May 5. The mosaics show a region of disturbance in the core that advances from a co-rotating longitude of $\lambda = 150^\circ$ (Fig. 1a) to one of $\lambda = 226^\circ$ (Fig. 1h) 133 days later. Furthermore, the disturbance is associated with a sequence of outward or inward extensions of material which we term 'jets'; these are initially radial but their slope (defined as the angle made with the tangential direction) decreases with time. Note that the individual jets always remain on one side of the core; they therefore represent a distribution in the semi-major axis ($a > a_F$ for outward jets and $a < a_F$ for inward jets, where $a_F = 140,223.7$ km is the length of the semi-major axis of the F ring's core⁷), rather than resulting from various eccentricities and longitudes of periape, as the latter would give rise to jets that oscillated about the core. The observed extent of major jets in the semi-major axis is ~ 500 km; if this separation were due to gravitational scattering, it would imply a body of diameter ~ 120 km, which would have been easily detected (see also Supple-

¹Astronomy Unit, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. ²C. E. de Saclay, Université Paris 7, L'Orme de Merisiers, 91191 Gif-sur-Yvette Cedex, France.

mentary Videos 1 and 2). An alternative and more compelling explanation is that jets (and hence spirals that form from them) are caused by physical collisions¹¹.

Material with a semi-major axis of $a_F + \Delta a$ would be expected to drift with respect to the core in such a way that after a time Δt its longitudinal drift relative to its initial position would be $\Delta\lambda = -(3/2)(n/a)\Delta a\Delta t$, where $n = 581.964^\circ$ per day is the mean motion of material in the core. Therefore, keplerian shear causes the slope of an initially radial jet to be $g = -(2/3)(a/n)/\Delta t$ after a time Δt : it is independent of Δa . Consequently, under certain assumptions, the gradient of a jet gives the time at which it was formed. After sufficient time, a radial jet will evolve into a spiral; indeed, the upper halves of the images in Fig. 1a–d display spiral structures exterior to the core, presumably arising from previous events.

It has been suggested¹¹ that S/2004 S 6 is a candidate for the object causing the collisional formation of the spiral structures observed by Cassini in mosaics taken in 2004 and 2005. The open circles in Fig. 1

show the predicted locations of S/2004 S 6 based on a new numerical fit to detections from 2004 October to 2005 June (see Supplementary Table 2). Another object, F07090QB, was detected and tracked in the 2007 March, April and May sequences (Fig. 1f–h), and its separately calculated orbital elements bear a striking resemblance to those of S/2004 S 6 (see Supplementary Table 2). The predicted and actual locations of F07090QB appear as open squares in Fig. 1a–e and Fig. 1f–h, respectively. We believe that this object is either S/2004 S 6 or some fragment of it, perhaps perturbed by encounters with material in the core. A more detailed analysis of these orbits will be presented elsewhere. However, the predicted or actual longitude of each object is clearly associated with the progress of the disturbed region, which is consistent with the tabulated differences in mean motion between the disturbing agents and the F-ring core given in Supplementary Table 2. We note that by early 2007 the difference in the longitudes of periape between the orbits of S/2004 S 6 and the F-ring core was $\sim 159^\circ$ and is currently increasing at a rate of 1.6° y^{-1} ;

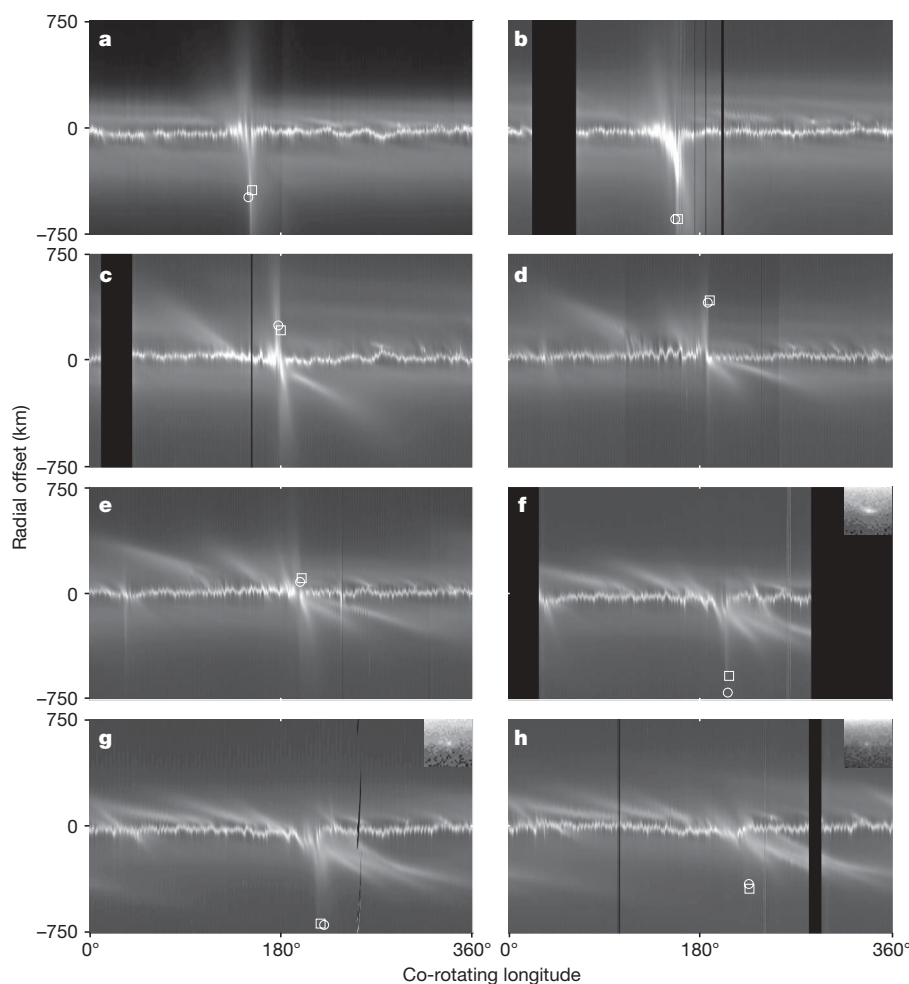


Figure 1 | Sequence of 360° mosaics of the F ring showing evidence for a series of collisional events between December 2006 and May 2007. In each sequence the mosaic is constructed by re-projecting individual images from Cassini's narrow-angle camera into a frame in which the horizontal axis is the longitude at the same epoch (12:00 UTC on 2007 January 1) and the vertical axis is the orbital radius, r , relative to the adopted orbit of the F ring⁷. Black areas indicate lack of longitudinal coverage and all projections are onto the equatorial plane. In each mosaic the circle marks the predicted location of S/2004 S 6 using the orbital elements listed in Supplementary Table 2. The square marks either the predicted (a–e) or actual (f–h) location of the new object using the orbital elements listed in Supplementary Table 2. The inserts in the upper right-hand corners of f–h are enhanced, re-oriented 30×30 -pixel crops of representative raw images showing the objects F07090QB, F07108QA and F07125QB, respectively (see Supplementary Table 2). a, Image

observation ISS_036RF_FMOVIE001_VIMS, consisting of 127 images obtained 2006 December 23–24. b, Image observation ISS_036RF_FMOVIE002_VIMS, consisting of 130 images obtained 2007 January 5–6. c, Image observation ISS_039RF_FMOVIE002_VIMS, consisting of 124 images obtained 2007 February 10–11. d, Image observation ISS_039RF_FMOVIE001_VIMS, consisting of 143 images obtained 2007 February 27. e, Image observation ISS_041RF_FMOVIE002_VIMS, consisting of 169 images obtained 2007 March 17. f, Image observation ISS_041RF_FMOVIE001_VIMS, consisting of 107 images obtained 2007 March 31. g, Image observation ISS_043RF_FMOVIE001_VIMS, consisting of 94 images obtained 2007 April 18. h, Image observation ISS_044RF_FMOVIE001_VIMS, consisting of 137 images obtained 2007 May 5. Further information about the observations is given in Supplementary Table 1.

a typical encounter velocity with the core is $\sim 30 \text{ m s}^{-1}$, in agreement with the estimated collisional velocity¹¹.

Figure 2 shows a comparison between an enlarged version of Fig. 1g and the results of a numerical simulation of jets evolved from initial radial distributions. The observed features in Fig. 2a that correspond to our modelled jets O₁, O₂, O₃, I₁, I₂, I₃ and I₆ (Fig. 2b) are consistent with collisions between the F-ring core (or objects orbiting close to it) and S/2004 S 6. However, the smaller jet pairs, O₄ and I₄ and O₅ and I₅, have different characteristics and locations that are inconsistent with collisions with S/2004 S 6; these jets may be attributed to collisions with additional objects. Furthermore, Fig. 2a shows three smaller jets with shallow gradients, and an associated bright feature, just above the core at co-rotating longitudes 250°, 295° and 335°; these can also be seen in Fig. 1, moving from right to left. All seem to be associated with objects with semi-major axes $\sim 50 \text{ km}$ greater than that of the F ring (see below).

If every close approach between S/2004 S 6 and the core were to lead to a collision, we would expect ~ 200 jets to be detectable in Fig. 1.

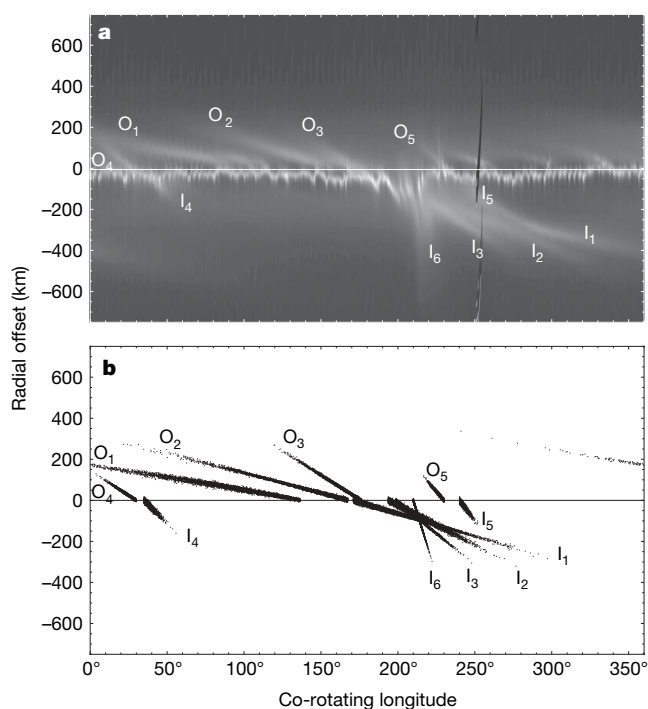


Figure 2 | Comparison of a re-projected mosaic with the results of a numerical simulation of jet formation and evolution. **a**, Re-projected F-ring mosaic derived from observation ISS_043RF_FMOVIE001_VIMS (see Fig. 1g). **b**, Results of a numerical simulation of jets evolved to the date of the mosaic, 2007 April 18. In each case the data are shown in the standard co-rotating longitude system used in Fig. 1. For the simulation, outward jets of particles were created on 2006 December 18 (O₁, co-rotating longitude $\lambda = 137^\circ$), 2007 January 24 (O₂, $\lambda = 168^\circ$), March 15 (O₃, $\lambda = 176^\circ$), March 16 (O₄, $\lambda = 30^\circ$) and March 30 (O₅, $\lambda = 230^\circ$), and inward jets on 2007 February 6 (I₁, $\lambda = 171^\circ$), March 7 (I₂, $\lambda = 194^\circ$), March 22 (I₃, $\lambda = 199^\circ$), March 27 (I₄, $\lambda = 35^\circ$), April 1 (I₅, $\lambda = 240^\circ$) and April 11 (I₆, $\lambda = 210^\circ$). Between 2,000 and 8,000 particles were used to simulate individual events. The particles in each jet were given gaussian distributions in a , e and ϖ relative to the F-ring core model, with standard deviations in Δa of either 75 km for large jets or 35 km for smaller jets (half-gaussian distributions with positive values for outward jets and negative for inward jets), in Δe of 5×10^{-5} and in $\Delta \varpi$ of 0.5° . The initial longitude was set to the value at which the jet was created. The particles were allowed to precess under the effects of Saturn's gravitational harmonics J_2 , J_4 and J_6 and were assumed to orbit at a mean motion determined by their initial semi-major axes. Their resulting radial locations with respect to the F-ring core and associated co-rotating longitudes are plotted. Note that the formation times and initial longitudes in **b** were deliberately chosen to match the features in **a**; consequently, the timing should be considered as representative rather than exact.

However, Figs 1 and 2a demonstrate that the core is clearly non-uniform in nature; without detailed knowledge of its structure and time variation, as well as the orbits of all nearby objects, the outcome of any given encounter cannot be predicted. After a sufficient time, differential precession causes the gradients of the jets with respect to the core to undergo variations on orbital timescales. However, if the core has sufficient mass, then secular perturbations²⁰ may induce the jet material to maintain approximate apse alignment with the core²¹. Consequently, detailed observations of the formation and evolution of jets may constrain the F ring's mass. Clearly the orbit of S/2004 S 6 is also subject to the combined secular effects of Prometheus and any mass contained within the F ring^{22,23} (see Supplementary Figs 2 and 3). This may help to explain why the mosaic seen in Fig. 2a closely resembles those seen two years previously (see Supplementary Fig. 1).

Embedded objects

The existence of sheared jets of material produced by collisions between objects and the F-ring core has an unexpected consequence: it provides tracer material that allows us to detect the gravitational (or

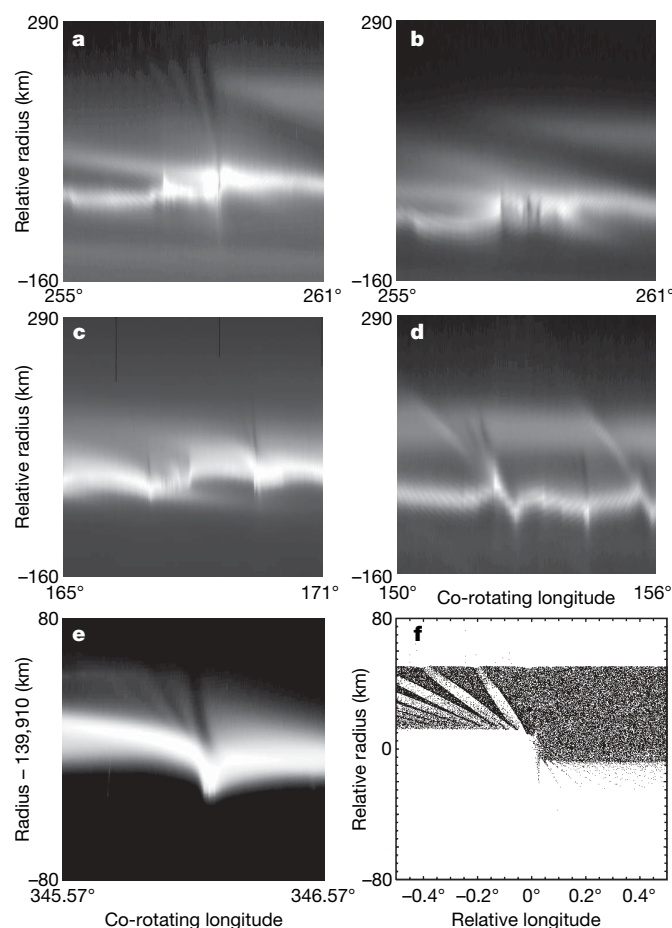


Figure 3 | Evidence for embedded satellites on eccentric orbits near the core of the F ring. **a**, Re-projected 6° section of the F-ring mosaic derived from image N1538169712, obtained 2006 September 28 at 20:50 UTC, showing the existence of a fan associated with the bright clump in the core. **b**, The same 6° longitudinal range as in **a**, but for the image N1538870080, obtained 2006 October 6 at 23:23 UTC. **c**, Fans in a 6° re-projected section of image N1552819237, obtained 2007 March 17 at 10:07 UTC. **d**, Fans in a 6° re-projected section of image N1557039476, obtained 2007 May 5 at 06:24 UTC. **e**, 1° re-projected section of image N1537876778, obtained 2006 September 25 at 11:28 UTC, showing evidence of a fan. **f**, Numerical simulation of the gravitational effect of a satellite (radius $\sim 5 \text{ km}$, density 0.4 g cm^{-3} , $\Delta e = 1.8 \times 10^{-4}$) embedded at the inner edge of a 50-km-wide F ring composed of test particles; note that the scale is the same as in **e**. See also Supplementary Video 3 and Supplementary Fig. 4.

even collisional) signature of embedded satellites. Figure 3a shows an example of a feature consisting of a sequence of sheared channels with a common point of intersection; we refer to this as a 'fan'. In this case the fan is visible in jet material lying exterior to the F-ring core, in an image taken before the sequences shown in Fig. 1. The channels point to a bright section of the ring with a shallow-sloped, outward, sheared jet going off to the left and a broad jet of material at the upper right. When the same region was viewed 8.1 days later (Fig. 3b), the main structures are still visible but, although there is no fan, dark channels are present in the F ring. Smaller fans are visible in Fig. 3c (a section from Fig. 1e) and Fig. 3d (a section from Fig. 1h), on the same scale as Fig. 3a, b. In Fig. 3d the diagonal features are streamers—channels associated with the previous passage of Prometheus, and the left-hand feature has a fan above it; note the second, smaller fan located between the two streamer—channel features.

Fans are produced by the gravitational effect of an embedded satellite on an orbit that has a non-zero relative eccentricity with respect to the adjacent material. If the relative eccentricity is zero (that is, the satellite has the same orbital eccentricity and periape as the F ring), then the satellite produces a stationary pattern on the ring material that approaches it. This is identical to the mechanism that is known to generate the propeller features detected in Saturn's A ring²⁴. If the perturbing satellite has a semi-major axis at the inner edge of a ring of material, only the upper half of a propeller is produced. According to our numerical models, once a relative eccentricity is introduced, the satellite's movement along the ring causes a fan to be produced, with the process continuing as long as material is available to be perturbed. Furthermore, the models demonstrate that the fan goes through a cycle, appearing and disappearing on an orbital timescale, thereby explaining Fig. 3a, b.

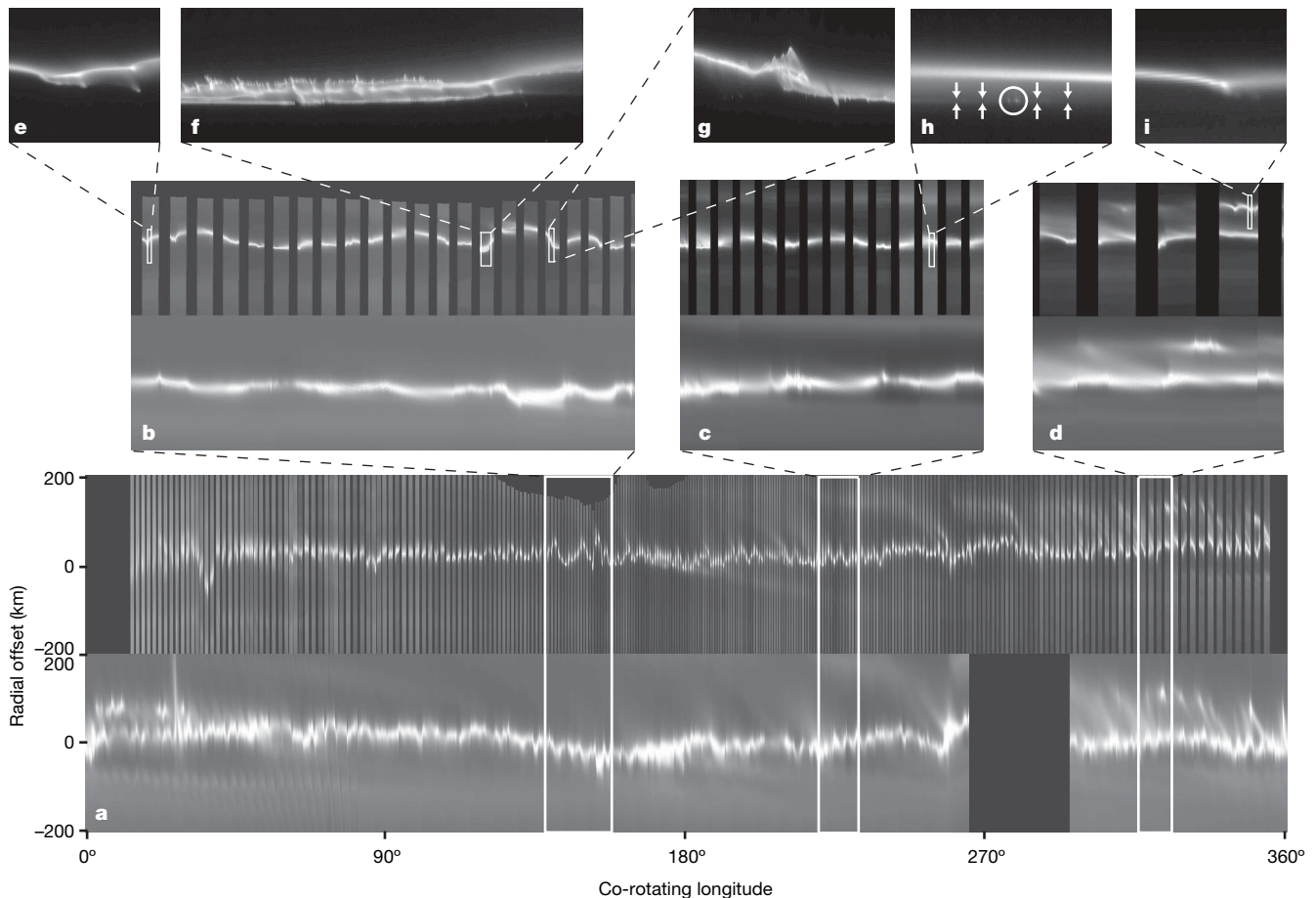


Figure 4 | Dynamical structures on a variety of scales in the F ring.

a, Comparison of two 360° mosaics of the F ring; the upper and lower mosaics are respectively derived from sequence ISS_029RI_AZSCNLOPH001_PRIME, obtained 2006 September 25 between 11:18 and 21:07 UTC (consisting of 269 non-overlapping images), and sequence ISS_029RF_FMOVIE001_VIMS, obtained 2006 September 28–29 between 20:32 and 10:17 UTC (consisting of 93 overlapping images). Each mosaic is constructed from re-projected images covering a 400-km radial range centred on the calculated radius of the F ring's core and with a co-rotating longitude, λ , shifted to a common epoch (12:00 UTC, 2007 January 1, as in Fig. 1). We note that the pointing of images in the upper mosaic for $\lambda \lesssim 100^\circ$ is only approximate, owing to a lack of stars in the original images. The highest resolution in the upper mosaics is ~ 1 km per pixel in both the radial and longitudinal directions. The extended radial feature at $\lambda = 26^\circ$ in the lower mosaic is associated with an object that has also been detected in the upper mosaic. **b–d**, Enlarged sections of the mosaics in **a** showing the regions with $138^\circ \leq \lambda \leq 158^\circ$ (**b**), $220^\circ \leq \lambda \leq 232^\circ$ (**c**) and

$316^\circ \leq \lambda \leq 326^\circ$ (**d**); all sections are shown on a common scale. The corresponding inertial longitudes covered are $136^\circ < \lambda_1 < 138^\circ$, $264^\circ < \lambda_2 < 271^\circ$ (**b**); $139^\circ < \lambda_1 < 141^\circ$, $263^\circ < \lambda_2 < 270^\circ$ (**c**); and $166^\circ < \lambda_1 < 172^\circ$, $265^\circ < \lambda_2 < 273^\circ$ (**d**), where the subscripts '1' and '2' denote the upper and lower mosaics, respectively. **e–g**, Enlarged sections of the mosaics in **b**, showing the respective regions with $138.97^\circ \leq \lambda \leq 139.12^\circ$ ($139,950 \text{ km} \leq r \leq 140,050 \text{ km}$), derived from image N1537900950; $152.32^\circ \leq \lambda \leq 152.72^\circ$ ($139,940 \text{ km} \leq r \leq 140,040 \text{ km}$), image N1537899083; and $154.99^\circ \leq \lambda \leq 155.19^\circ$ ($139,950 \text{ km} \leq r \leq 140,050 \text{ km}$), image N1537898708. **h**, Enlarged section of the mosaic in **c**, showing the region with $230.27^\circ \leq \lambda \leq 230.47^\circ$ ($139,950 \text{ km} \leq r \leq 140,030 \text{ km}$), image N1537888059; note the ~ 1 -km-wide linear feature (indicated by the arrows) below the bright component and the 2–3 embedded objects in it (within the circle). **i**, Enlarged section of the upper mosaic in **d** showing the region with $324.94^\circ \leq \lambda \leq 325.09^\circ$ ($139,950 \text{ km} \leq r \leq 140,050 \text{ km}$), image N1537878429. We note that the scales in **e–i** are the same. See Supplementary Figs 5–9 for higher resolution versions of **e–h**.

As a representative example of the mechanism, Fig. 3e, f directly compares a fan image to a frame from a numerical simulation of a 50-km-wide ring perturbed by an embedded satellite of ~ 5 km radius on an eccentric orbit. In our simulations, the maximum longitudinal width of the channels that make up a fan is $\sim 10R_H$, where $R_H = a(\mu/3)^{1/3}$ is the radius of the satellite's Hill sphere, μ being the ratio of the satellite's mass to that of the planet. Given that the widest channel in Fig. 3e is ~ 100 km wide, this would imply that it is caused by a satellite of radius ~ 7 km, assuming it has a density comparable to that of Prometheus; this is in good agreement with the value used in the numerical simulation. A similar analysis would suggest that the object responsible for the fan visible in Fig. 3a has a radius of ~ 35 km; however, the presence of two or three dark bands at this location in Fig. 3b suggests that multiple objects may be responsible. The calculated radii for the objects causing the fans seen in Fig. 3c, d are correspondingly smaller.

Figure 4a shows two mosaics of the region within 200 km of the F-ring core, using re-projected images from 2006 September. The Prometheus channel signature is evident in the lower mosaic for $\lambda < 81^\circ$ across the whole radial region (see also the patterns at $\lambda < 75^\circ$ in Fig. 2a), as are various jets. The sheared channel¹⁰ patterns are a familiar feature of the F-ring region. However, there are several hundred additional, localized distortions in the F-ring core, examples of which can be seen in Fig. 4b–d. Our numerical models suggest that the effect of keplerian shear would make the channels undetectable within $\sim 100^\circ$ (or ~ 20 days) of a Prometheus conjunction, yet Fig. 4a–d shows that the core still retains a 'memory' of the last encounter. These localized distortions are time variable; for example, Fig. 4b displays a region of the core covering the same range of co-rotating longitude but in which the inertial longitude differs by $\sim 130^\circ$ between the two mosaics. We note the two small sections of the ring that point downwards in the upper mosaic and upwards in the lower mosaic. This reflective symmetry is apparent throughout Fig. 4b and, to a lesser extent, in Fig. 4c, d; the phenomenon was first noted in pre-orbit insertion images (see fig. 8c, d of ref. 8) and is indicative of ring material with a small, induced eccentricity and/or periape relative to the core. There is a tendency for the distortions always to point in the same direction, suggestive of some coherent perturbation similar to the streamer–channel phenomenon produced by Prometheus.

Any given location in the F-ring core has a conjunction with Prometheus every 67.65 days, but the resulting maximum perturbation to the core depends on the relative alignment of the two orbits; because of differential precession this follows a ~ 19 -year cycle with the next anti-alignment due in November 2009 (ref. 25). The maximum absolute values of the perturbation in a at alignment, at the orientation when the periapees differ by 90° and at anti-alignment are < 0.5 km, 4 km and 19 km, respectively. The corresponding changes in eccentricity, e , and longitude of periape, ϖ , are 2, 6 and 13×10^{-5} , and 0.5° , 1.5° and 4° , respectively. Although the resulting distortions to a ring of test particles shear away, the same would not be true for the orbits of small satellites at the locations of Prometheus conjunctions. We conclude that the apparent memory of the core region is caused by Prometheus perturbing masses that lie in or close to the F-ring core, or perhaps core material itself if it is sufficiently massive. Once the orbital elements of these objects are excited, only the self-gravity of the ring and/or collisions can dampen them. Consequently they will, in turn, perturb surrounding ring material or collisional debris, and leave a gravitational signature (see Supplementary Video 4).

Figure 4e–i presents selected regions of the high-resolution mosaic on a common scale. Although not always easy to interpret without temporal coverage, each image shows evidence for the existence of small (diameter < 1 km) objects in the F-ring region. Such objects may have associated dust envelopes hindering direct imaging of their surfaces. A common feature of Fig. 4e–g is that the location of each presumed object is correlated with a corresponding feature in the bright core, implying a dynamical connection. The downwards-pointing features in the jet in the upper right-hand region of Fig. 4i

are not seen in the core below, showing that the perturbation is localized; this demonstrates that such perturbed objects can exist at least 50 km from the core. This is the feature seen at $\lambda = 250^\circ$ in Fig. 2a (see above).

A narrow ring component

In images from the high-resolution observation ISS_029RI_AZSCNLOPH001_PRIME (upper mosaics in Fig. 4) we have identified a non-continuous, narrow (~ 1 -km-radial-width) ring component located ~ 15 – 50 km radially inward from the F-ring core in 45 images with $90^\circ < \lambda < 270^\circ$, a range that includes the highest resolution images; in each case the phase angle is $< 90^\circ$. Eight of these images show evidence of embedded objects (diameter ~ 1 km) associated with the narrow component. For example, Fig. 4h contains two or three embedded objects. This and additional images showing the narrow component are shown in Supplementary Figs 9 and 10. A preliminary fit to detections of this component strongly supports the suggestion that it has an orbit that is closely approximated (root-mean-square residuals < 1.34 pixels) by the orbital solution⁷ for the F ring's core derived from occultation data; in fact, it constitutes a better fit than that based on Cassini observations of the F-ring core. We note that all the re-projected images in Figs 1, 2 and 4 calculate radius with respect to the same orbital model of the core⁷, and all clearly show deviations of ~ 50 km. The discovery of such a narrow component would be consistent with the Voyager radio science^{4,5} and photopolarimeter⁶ occultation data but inconsistent with the broad feature present in the data from the Cassini Ultraviolet Imaging Spectrograph and Visual and Infrared Mapping Spectrometer¹⁵; it is probable that the latter instruments are detecting the bright F-ring core (width ~ 50 km) visible in all the Cassini images.

Discussion

The existence of a population of moonlets colliding with the F-ring core on a regular basis makes this region unique in the Solar System as a location where the effects of on-going collisions at $\sim 30 \text{ m s}^{-1}$ can be studied *in situ*. Direct observations will serve to constrain realistic collisional modelling and lead to an improved understanding of the role of such processes in a variety of locations and epochs.

We attribute the ability of the F ring to retain a memory of previous conjunctions with Prometheus to the existence of perturbed, ~ 1 -km-diameter objects associated with the core; larger objects on more eccentric orbits produce the fan structures we have identified in the Cassini images. Therefore, Prometheus plays two roles in perturbing the F ring: its direct perturbations imprint the streamer–channel structures across the whole region as it passes through every 67.65 days¹⁰, whereas the same perturbations excite embedded objects that proceed to exert their own gravitational effect on surrounding material. A probable consequence is that Prometheus itself must be continually struck by material from this region. Future high-resolution imaging of the moon could reveal the effects of recent, direct collisions, perhaps in the form of an asymmetry between Prometheus's leading and trailing hemispheres.

It is difficult to understand how the observed ~ 1 -km-wide ring component seen in some of the highest resolution images can survive in such a chaotic environment. Nonetheless, the evidence suggests not only that it does, but also that it even maintains enough integrity to precess uniformly; the only obvious mechanisms to prevent its destruction are self-gravity and collisions. For this reason, the mass of the F ring perhaps resides in this narrow component rather than the broad, bright, dusty²⁶ core seen in every image. This would imply that the narrow component be gravitationally unstable, thereby providing a natural place for moonlet formation. A combination of additional high-resolution imaging and detailed numerical modelling (including the effects of collisions and self-gravity) will be necessary if progress is to be made in this area.

A more general question is how were such a ring and its retinue of objects formed. One possibility is that the system is young ($\sim 10^6$ years old) and the result of the collisional disruption of a moon by an impact event. This could explain the relatively large dispersion of its constituent objects in contrast to the comparative regularity of Saturn's main ring system adjacent to it. The F ring's location also means that its structure could be the result of a continual battle between accretion and disruption¹⁷ on a variety of scales.

Received 14 February; accepted 7 April 2008.

- Gehrels, T. *et al.* Imaging photopolarimeter on Pioneer Saturn. *Science* **207**, 434–439 (1980).
- Smith, B. A. *et al.* Encounter with Saturn: Voyager 1 imaging science results. *Science* **212**, 163–191 (1981).
- Smith, B. A. *et al.* A new look at the Saturn system: The Voyager 2 images. *Science* **215**, 504–537 (1982).
- Tyler, G. L., Marouf, E. A., Simpson, R. A., Zebker, H. A. & Eshelman, V. R. The microwave opacity of Saturn's rings at wavelengths of 3.6 and 13 cm from Voyager 1 radio occultations. *Icarus* **54**, 160–188 (1983).
- Marouf, E. A., Tyler, G. L. & Rosen, P. M. Profiling Saturn's rings by radio occultation. *Icarus* **68**, 120–166 (1986).
- Lane, A. L. *et al.* Photopolarimetry from Voyager 2: Preliminary results on Saturn, Titan and the rings. *Science* **215**, 537–543 (1982).
- Bosh, A. S., Olkin, C. B., French, R. G. & Nicholson, P. D. Saturn's F ring: Kinematics and particle sizes from stellar occultation studies. *Icarus* **157**, 57–75 (2002).
- Porco, C. C. *et al.* Cassini imaging science: Initial results on Saturn's rings and small satellites. *Science* **307**, 1226–1236 (2005).
- Giulietti Winter, S. M., Murray, C. D. & Gordon, M. Perturbations to Saturn's F ring strands at their closest approach to Prometheus. *Planet. Space Sci.* **48**, 817–827 (2000).
- Murray, C. D. *et al.* How Prometheus creates structure in Saturn's F ring. *Nature* **437**, 1326–1329 (2005).
- Charnoz, S. *et al.* Cassini discovers a kinematic spiral ring around Saturn. *Science* **310**, 1300–1304 (2005).
- Porco, C. C. Satellites and rings of Saturn. *IAU Circ.* **8432** (2004).
- Showalter, M. R. Detection of centimeter-sized meteoroid impact events in Saturn's F ring. *Science* **282**, 1099–1102 (1998).
- Showalter, M. R. Disentangling Saturn's F ring. I. Clump orbits and lifetimes. *Icarus* **171**, 356–371 (2004).
- Esposito, L. W., Meinke, B. K., Colwell, J. E., Nicholson, P. D. & Hedman, M. M. Moonlets and clumps in Saturn's F ring. *Icarus* **194**, 278–289 (2008).
- Cuzzi, J. N. & Burns, J. A. Charged particle depletion surrounding Saturn's F ring: Evidence for a moonlet belt? *Icarus* **74**, 284–324 (1988).
- Barbara, J. M. & Esposito, L. W. Moonlet collisions and the effects of tidally modified accretion in Saturn's F ring. *Icarus* **160**, 161–171 (2002).
- Poulet, F., Sicardy, B., Nicholson, P. D., Karkoschka, E. & Caldwell, J. Saturn's ring-plane crossing of August and November 1995: A model for the new F ring objects. *Icarus* **144**, 135–148 (2000).
- Porco, C. C. *et al.* Cassini imaging science: Instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004).
- Murray, C. D. & Dermott, S. F. *Solar System Dynamics* Ch. 7 (Cambridge University Press, Cambridge, UK, 1999).
- Dermott, S. F. & Murray, C. D. Origin of the eccentricity gradient and the apse alignment of the epsilon ring of Uranus. *Icarus* **43**, 338–349 (1980).
- Murray, C. D. *et al.* Saturn's F ring and its retinue. *Bull. Am. Astron. Soc.* **37**, 3 (Annual DPS meeting), abstr. 64.05 (2005).
- Spitale, J. N., Jacobson, R. A., Porco, C. C. & Owen, W. M. The orbits of Saturn's small satellites derived from combined historic and Cassini imaging observations. *Astron. J.* **132**, 692–710 (2006).
- Tiscareno, M. S. *et al.* 100-metre-diameter moonlets in Saturn's A ring from observations of 'propeller' structures. *Nature* **440**, 648–650 (2006).
- Borderies, N. & Goldreich, P. The variations in eccentricity and apse precession rate of a narrow ring perturbed by a close satellite. *Icarus* **53**, 84–89 (1983).
- Showalter, M. R., Pollack, J. B., Ockert, M. E., Doyle, L. R. & Dalton, J. B. A photometric study of Saturn's F ring. *Icarus* **100**, 394–411 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements C.D.M., K.B., N.J.C., M.W.E. and G.A.W. are grateful to the UK Science and Technology Facilities Council for financial support. The success of the FMOVIE observations described in this paper is due to the invaluable assistance of M. Showalter and the Cassini VIMS team. The authors also wish to thank R. Nelson, D. McNeil, J. Burns and our other colleagues on the Cassini project for numerous discussions.

Author Contributions C.D.M. led the research, worked on image analysis and carried out the jet simulations. K.B. and M.W.E. wrote database and image analysis software, and contributed to image sequence design. Satellite and ring orbit analysis was done by N.J.C., as were the simulations of S/2004 S 6 encountering the F ring. Simulations of planar ring–satellite interactions were done by G.A.W. C.D.M., K.B., N.J.C., M.W.E., G.A.W. and S.C. all contributed to the theories discussed.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.D.M. (c.d.murray@qmul.ac.uk).

The branching programme of mouse lung development

Ross J. Metzger^{1†}, Ophir D. Klein^{2†}, Gail R. Martin² & Mark A. Krasnow¹

Mammalian lungs are branched networks containing thousands to millions of airways arrayed in intricate patterns that are crucial for respiration. How such trees are generated during development, and how the developmental patterning information is encoded, have long fascinated biologists and mathematicians. However, models have been limited by a lack of information on the normal sequence and pattern of branching events. Here we present the complete three-dimensional branching pattern and lineage of the mouse bronchial tree, reconstructed from an analysis of hundreds of developmental intermediates. The branching process is remarkably stereotyped and elegant: the tree is generated by three geometrically simple local modes of branching used in three different orders throughout the lung. We propose that each mode of branching is controlled by a genetically encoded subroutine, a series of local patterning and morphogenesis operations, which are themselves controlled by a more global master routine. We show that this hierarchical and modular programme is genetically tractable, and it is ideally suited to encoding and evolving the complex networks of the lung and other branched organs.

Many organs are composed of highly ramified tubular networks, each with a distinct architecture tailored to its physiological function. The bronchial tree of the human lung has more than 10^5 conducting and 10^7 respiratory airways arrayed in an intricate pattern crucial for oxygen flow^{1–4}. Classical studies of lung structure^{5–8} raise the question of how the information required to generate a tree of such complexity is biologically encoded⁹. Individually configuring thousands or millions of branches would require a tremendous amount of patterning information, far more than is biologically plausible, to specify when and where each branch forms during development, and the size, shape and direction of outgrowth of each branch. One possibility is that the process is not precisely controlled; for example, if branching occurs randomly to fill available space. Another is that control is precise but coding is simplified by repeated use of a branching mechanism, as in Mandelbrot's fractal model and other elegant algorithms^{10–17}.

Even with these attractive models and recent progress in identifying lung development genes¹⁸, understanding of the programme that directs branching remains rudimentary. This is largely due to the complexity of the bronchial tree, which makes it difficult to follow branching dynamics beyond the earliest events^{19–21}. Although branching of the lung and other organs can occur in culture^{22–25}, it is unlikely that these recapitulate the full pattern. Here we describe the complete *in vivo* pattern of branching and branch lineage of the mouse bronchial tree, and show that it is generated using three geometrically distinct local modes of branching coupled in three different sequences.

The branch lineage of the mouse bronchial tree

The bronchial tree develops by branching of the airway epithelium into surrounding mesenchyme. Although the process cannot be visualized in living embryos with current techniques, we reasoned we could reconstruct the branching sequence from fixed specimens, provided that the process is stereotyped. An immunostaining

procedure was developed to visualize the full three-dimensional structure of the bronchial tree in fixed lungs (Fig. 1a). Examination of hundreds of wild-type CD1 specimens collected between embryonic day (E)11 and E15 revealed that the branching pattern is remarkably stereotyped. This allowed us to reconstruct the sequence of events—where, when and in what order branches form—from finely staged specimens (Fig. 1b). This information was used to construct a lineage diagram representing the developmental history of the ~5,000 branches of the bronchial tree (Fig. 1c, d and Supplementary Fig. 1). We found that there are three branching modes used repeatedly throughout the lung, which we call domain branching, planar bifurcation and orthogonal bifurcation.

Domain branching

In domain branching, daughter branches form in rows ('domains') at different positions around the circumference of the parent branch, like the rows of bristles on a bottle brush (Fig. 2f). In the left primary bronchus (L) lineage, the first secondary branch (L.L1, abbreviated L1) buds off the lateral aspect of the founder branch L late on E11 (Fig. 1b). Over the next two days, additional branches sprout distal to L1, creating a row of lateral secondary branches numbered in the proximal–distal sequence in which they form (L1, L2, and so on; Figs 1b and 2a). As these sprout, another row begins to form along the dorsal surface of L. The first dorsal branch (D1) buds just distal to the level of L1, and others bud sequentially in proximal-to-distal order (Fig. 2a, b). As this domain develops, a third row begins to sprout from the medial surface of L, and then a fourth from the ventral surface (Fig. 2b, c). This ventral domain often consisted of just a single branch (V1) located distally, and sometimes there were none. Although rudimentary, this is a *bona fide* domain because we found rare wild-type variants and a mutant that form more complete rows (see below).

Secondary branches off RCd (the distal portion of the R primary branch) also arise by domain branching, beginning with a row of

¹Department of Biochemistry and HHMI, Stanford University School of Medicine, Stanford, California 94305-5307, USA. ²Department of Anatomy and Program in Developmental Biology, School of Medicine, University of California at San Francisco, San Francisco, California 94158-2324, USA. †Present addresses: Department of Anatomy, School of Medicine, University of California at San Francisco, California 94158-2517, USA (R.J.M.); Departments of Orofacial Sciences and Pediatrics, and Institute of Human Genetics, Schools of Dentistry and Medicine, University of California at San Francisco, San Francisco, California 94143-0442, USA (O.D.K.).

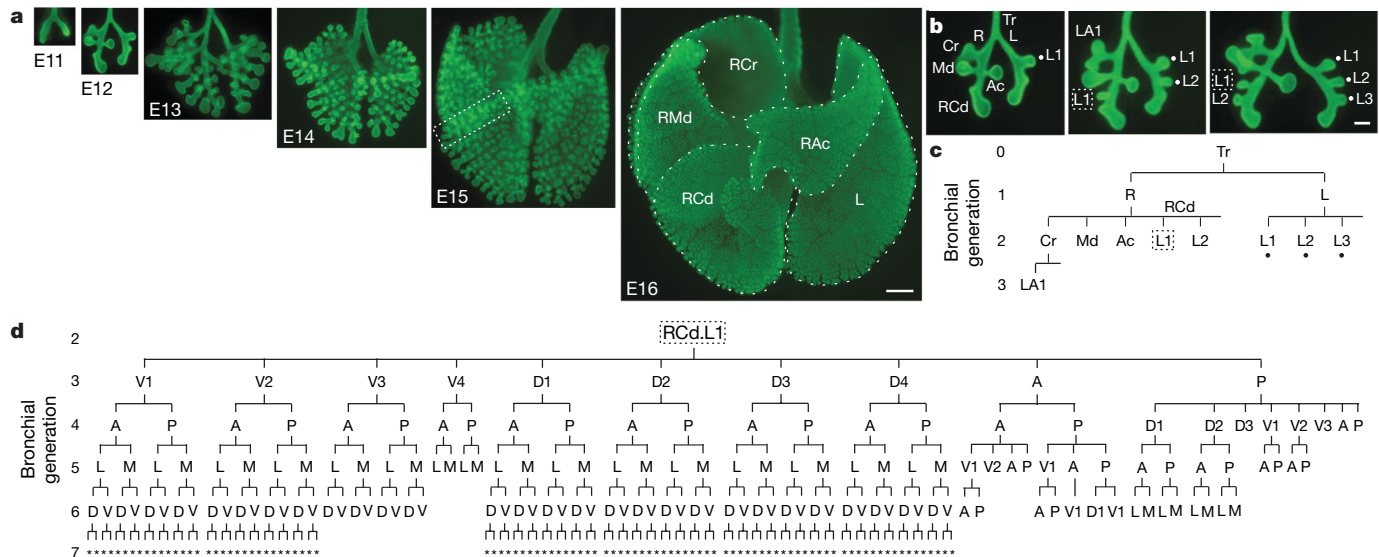


Figure 1 | Branching morphogenesis of the mouse bronchial tree. **a**, Whole-mount lungs (ventral view) at the embryonic day indicated immunostained for E-cadherin (green) to show the airway epithelium. Dotted lines show the right cranial (RCr), right middle (RMd), accessory (RAc), right caudal (RCd) and left (L) lobes. Scale bar, 500 μ m. **b**, Reconstructing branching dynamics using three E12 specimens \sim 3 h apart in age. Lateral secondary branches L1–3 (dots in **b**, **c**) sprout in a proximal-to-distal order from the left (L)

primary branch, as do the lateral secondary branches L1 (box in **b**, **c**) and L2 from the distal (RCd) portion of the right (R) primary branch. Scale bar, 200 μ m. **c**, Branch lineage diagram for the oldest lung in **b**. Branch names indicate the lineage, for example, RCd.L1 is first lateral secondary branch off RCd. **d**, Lineage diagram of RCd.L1 showing 250 descendant branches at E15 (box in **a**). A, anterior; D, dorsal; L, lateral; M, medial; P, posterior; V, ventral; asterisk, orientation can vary.

lateral branches (Fig. 1b). The spacing of branches in each row and the order in which rows trigger (Fig. 2c) are the same as in the L lineage, but the proximal–distal positions at which rows initiate and the

number of branches in each row are not. For example, the first dorsal branch (RCd.D1) forms proximal to the first lateral branch (RCd.L1), whereas in the L lineage the first dorsal branch (L.D1)

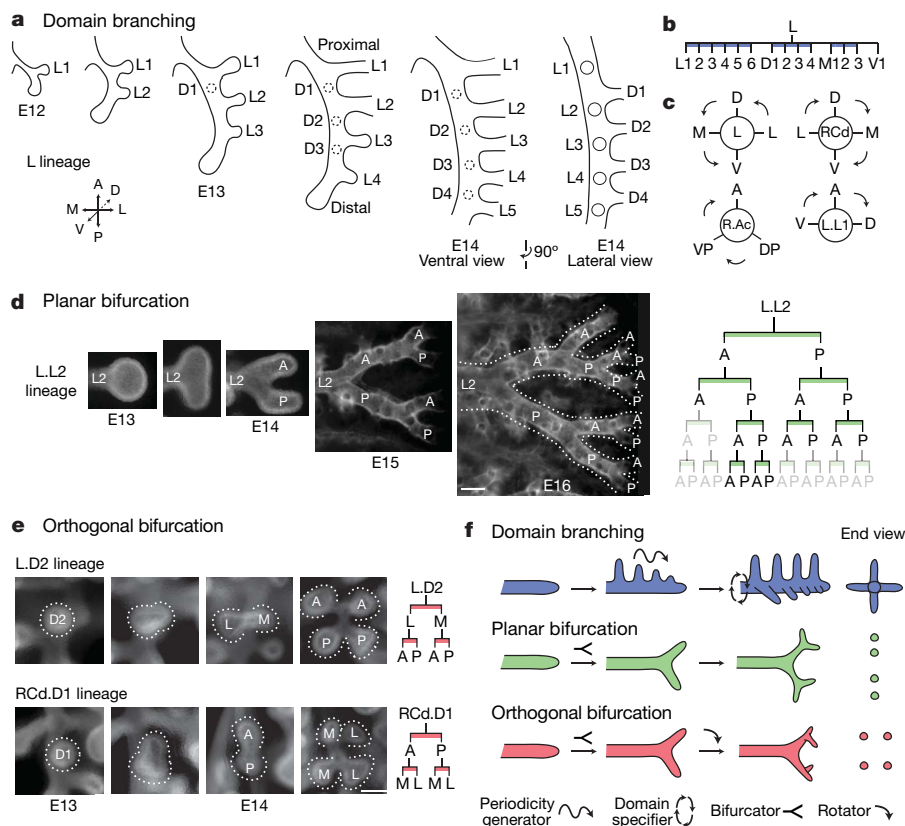


Figure 2 | Branching modes in lung development.

a–c, Domain branching. **a**, Schematics of lateral and dorsal secondary branches budding from L. Lateral secondary branches (L1–5) bud in proximal-to-distal order. Proximal-to-distal branching begins again in the second domain (projecting into the plane of the figure) to form a row of dorsal secondary branches (D1–4, dashed circles). Right panel, E14 schematic rotated 90° to show dorsal branches. **b**, Lineage diagram of secondary branches from L. Branches form in four domains: lateral (L), dorsal (D), medial (M) and ventral (V), indicated by blue bars. **c**, Schematic cross sections through L and the three other branches indicated, showing positions of domains and the order (arrows) in which domains are used. **d**, Planar bifurcation. Ventral view of the branch L.L2 in a series of fixed specimens from E13 to E16, showing sequential bifurcations along the A–P axis. E15 and E16 specimens were stained with anti-smooth muscle α -actin to highlight early branch generations, which are surrounded by smooth muscle. Dotted lines outline bifurcations. Right panel, lineage of L.L2 descendants formed by planar bifurcation; branches not yet formed in the E16 specimen are in grey. Scale bar, 100 μ m. **e**, Orthogonal bifurcation. End-on (dorsal) views of branches indicated in a developmental series of E13 and E14 specimens. L.D2 bifurcates along the L–M axis, and its daughters along the A–P axis, whereas RCd.D1 bifurcates along the A–P axis and its daughters along the L–M axis. Scale bar, 100 μ m. **f**, Schematics of branching modes. The first bifurcation in a series is classified retrospectively based on the orientation of the subsequent bifurcation. Icons show patterning and morphogenesis operations inferred for each mode: proximal–distal periodicity generator, circumferential domain specifier, branch bifurcator, and bifurcation plane rotator.

Variability and errors in the branch lineage

Although much of the branching process is stereotyped, it is not invariant. There was local variability in the temporal sequence of branching so that some lineages get ahead of, or fall behind, other lineages (Supplementary Fig. 2). There was also spatial variability including subtle differences in register between branches in different domains of a parent branch, and relaxation in the absolute orientation of later rounds of orthogonal bifurcation. None of this temporal or subtle spatial variability altered branch lineage. However, we also found variants that did affect lineage, which we call branching 'errors'.

Errors were identified in specimens with branch patterns that could not be reconciled with the canonical lineage unless an anomalous branching event had occurred. For example, there were specimens in which a branch originated off the wrong parent branch, a 'branch displacement' error (Fig. 4a, b, top panels), or in which a branch was missing and daughter branches sprang directly from the grandparent ('skipping a generation'; Fig. 4c, d, top panels). Despite their inappropriate origins, such branches went on to branch in their usual manner (Fig. 4a–d, lower panels), demonstrating that the information controlling a branch's subsequent branching is not

encoded within the parental branch, and that continuation of the branching programme is not contingent on completion of each previous step. We also found lungs in which a founder branch and all its descendants were missing, such as the L.V1 and RCd.V1 lineages (Fig. 2b and Supplementary Fig. 1). In each of these 'optional' lineages, the founder forms by domain branching and is typically the only branch formed in the last domain used.

Some errors, such as branch displacement, were rare, occurring at <1% of all branching events scored. Others were more common, and some, like skipping a generations in the RCd.L1 lineage, approached the frequency of the 'normal' event. Errors were not randomly distributed but tended to occur at specific sites and times in lung development. A more limited analysis of five inbred strains (A/J, C3H/HeJ, C57BL/6J, DBA/2J and FVB/NJ) showed the same pattern of branching and the same types of errors and other variation as in the outbred CD1 strain, with one exception detailed below. Thus, errors are not due to genetic heterogeneity. Rather, they identify intrinsically imprecise steps in the branching programme.

Genetic control of branch lineage and pattern

To begin to elucidate the genetic basis of the branching programme, we investigated branch lineage and pattern in two mutants and in an inbred strain with airway patterning defects. In *inversus viscerum* (*iv*) mutants in the *Dnahc11* dynein heavy chain gene²⁶, left–right axis specification is randomized and in about half the animals the positions and gross structures of organs are reversed²⁷. We found that the lung-branching pattern and lineage was completely reversed in some mutant embryos (Supplementary Fig. 3), demonstrating that deployment and coupling of the branching modes is under global genetic control and that it is downstream of *Dnahc11* and the left–right asymmetry pathway.

In contrast to this global effect, null mutations in sprouty 2 (*Spry2*)²⁸, encoding a sprouty family receptor tyrosine kinase inhibitor^{29,30}, had local and subtle effects on branch pattern and lineage: there were extra branches in the ventral domains off L and RCd (Fig. 5a, b and Supplementary Fig. 4). The extra branches sprouted earlier and more proximally than the normal branches in these domains, expanding the domains towards the base of the parent branch. The ectopic branches formed additional generations, creating ectopic lineages indistinguishable from those of normal branches in the domain. Thus, *Spry2* restricts the number of branches in the two ventral domains, and in its absence normally non-branching regions along the parent branches acquire the branching identity of more distal regions.

One inbred strain (C57BL/6J) had a subtle defect in branch positioning: branches in the dorsal domains of L and RCd were shifted distally along the parent branch (Fig. 5c, d and Supplementary Fig. 5). Despite this shift, the subsequent branch pattern and lineage of the displaced branches were unperturbed. We named the phenotype 'shifty' and propose that the shifty locus encodes a modulator of the pathway that sets the proximal–distal register of domains.

Discussion

The branching pattern and lineage of the mouse bronchial tree reveals the logic of the lung branching programme. Three local modes of branching (Fig. 2f) are used in three different sequences in the developing lung (Fig. 3). Each branching mode (domain branching, planar bifurcation and orthogonal bifurcation) creates a different arrangement of branches (bottle-brush, planar array and rosette, respectively) and serves a specific function in lung design (scaffold, edge and surface/interior). The repeated use of these branching modes, along with a hierarchical control and coupling scheme, allows genetic encoding of the complex but stereotyped bronchial tree.

A formal model of the airway branching programme. All three branching modes are geometrically simple and easy to encode. We propose that each is controlled by a locally operative, genetic

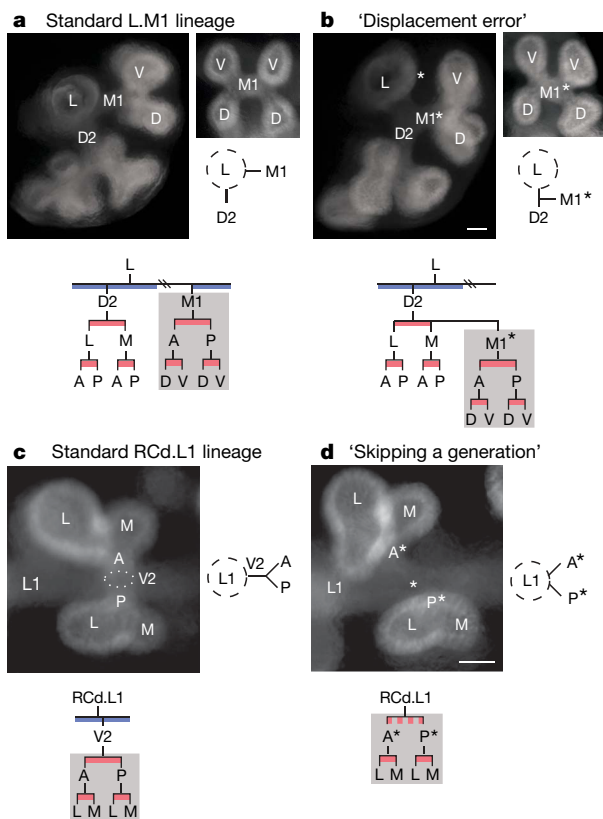


Figure 4 | Branching errors. **a, b, 'Displacement error'.** **a**, Standard arrangement of D2 and M1 secondary branches off L. Top-left panel, end-on (anterior) view of L and descendants. Top-right panel, end-on (medial) view of M1 and descendants; below, schematic of anterior view. Bottom panel, standard lineage of L, D2 and M1. Grey box, M1 lineage. **b**, Same views as **a** of lung in which M1* arises off D2, normally a sister branch. Scale bar (for **a** and **b**), 50 μ m. Bottom panel, M1* follows the normal M1 lineage (grey box). **c, d, 'Skipping a generation'.** **c**, Standard arrangement of V2 arising off RCd.L1. Top-left panel, end-on (ventral) view of V2 and its descendants, which form by orthogonal bifurcation. Top-right panel, schematic depicting side (lateral) view of V2 and its daughters. Bottom panel, standard lineage of RCd.L1 and V2. Grey box, descendants of V2. **d**, Same view (left) and schematic (right) of lung that skipped the V2 generation. Scale bar (for **c** and **d**), 50 μ m. Anterior (A*) and posterior (P*) branches sprout directly from RCd.L1, normally their grandparent. Bottom panel, descendants (grey box) of missing V2 follow the normal lineage.

subroutine—a series of discrete patterning and morphogenesis events (Figs 2f and 6). Domain branching requires a proximal–distal ‘periodicity generator’ that sets the timing and spacing of branches within a domain, and a circumferential ‘domain specifier’ that dictates the positions of domains around the parent branch and the order in which domains trigger. Planar and orthogonal bifurcation require a branch ‘bifurcator’, and orthogonal bifurcation requires a ‘rotator’ that reorients the bifurcation plane by 90° between events. All subroutines require a ‘branch generator’. Some of these steps may themselves be modular and shared among subroutines (Fig. 6).

Because the pattern of deployment of the branching modes is complex but stereotyped we infer that there is higher order, perhaps emergent, patterning information—‘the master routine’—which

calls subroutines at specific times and positions in the lung development programme. The three different sequences in which branching modes are combined (Fig. 3d) are each simple variants of a general coupling scheme (Fig. 6). Thus, the master routine need only encode the three variants, and specify where each is used. Coding is further simplified because most or all branches within a domain use the same scheme (Fig. 3d).

Although the sequences of subroutine use are rigidly specified, temporal variability in the programme (Supplementary Fig. 2) implies that the master routine does not function as a control centre that calls subroutines individually in a fixed global order. Rather, it appears to set the coupling scheme for each lineage early and then allow each lineage to proceed through its sequence independently. The programme is also regulative because branching continues normally after suffering errors (Fig. 4).

Because there are stereotyped local differences in the branching modes, such as the number of branches in a domain and the absolute orientation of orthogonal bifurcation, the master routine must also encode position-specific modifications in the subroutines, which we represent as local input parameters (P , Fig. 6). Setting these local parameters may be the most computationally intensive part of the programme.

Elucidating the genetic basis of the branching programme. A critical challenge ahead is to determine the genetic and molecular basis of the master routine, three subroutines and the local parameters. With the lineage in hand, functions can now be assigned with unprecedented precision to the dozens of extant lung development genes^{18,31–34}. We found that *Spry2* regulates the site of initiation and number of branches in specific domains ($P1$, Fig. 6), and shifty controls the proximal–distal register of entire domains ($P2$, Fig. 6). It will be particularly important to identify genes that underlie the periodicity generator, domain specifier, bifurcator and rotator, because they are central to the distinctive geometries of the branching modes and are likely to involve novel patterning processes.

Airway branching is one of many processes required to build a lung. Others include airway size control, airway cell differentiation,

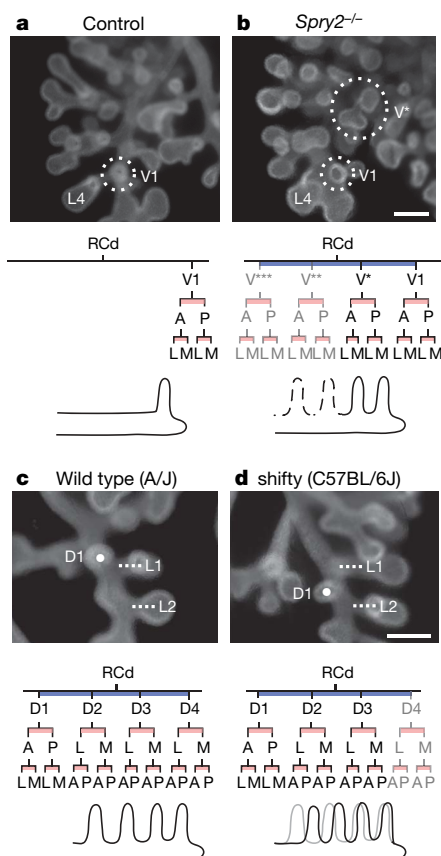


Figure 5 | Genetic control of branch pattern and lineage. **a, b**, Ectopic domain branching in *Spry2*^{-/-} mutants. **a**, The RCd lobe (ventral view) of an E12.5 control (*Spry2*^{+/+}) lung with a single secondary branch (V1, circled) off RCd, at the level of RCd.L4 (L4). Below, RCd.V1 lineage and schematic of RCd with a single ventral secondary branch (V1). **b**, Same view of an E12.5 *Spry2*^{-/-} lung showing the normal ventral secondary branch (V1) and an ectopic branch (V*) that forms earlier and proximal to V1. V* has already sprouted additional generations of branches. Below, lineage and schematic show V* plus additional ectopic ventral branches (V**, V***; dashed lines in schematic) seen in other *Spry2*^{-/-} lungs (Supplementary Fig. 4). Scale bar (for **a** and **b**), 200 μ m. **c, d**, Shifted domains in strain C57BL/6J. **c**, RCd lobe (dorsal view) of an E12.5 lung from the control strain A/J. The dorsal secondary branch RCd.D1 (D1; white dot) forms just proximal to the lateral secondary branch RCd.L1 (L1, white line). Below, lineage and schematic of RCd.D1 and other branches in the D domain. **d**, The same view of an E12.5 lung from a C57BL/6J mouse with the shifty phenotype. RCd.D1 forms distal to RCd.L1. In shifty lungs, the entire dorsal domain (black line in schematic) is shifted distally along RCd relative to the wild type (grey line in schematic), but the lineage is unaffected (except when there is a full unit shift and the RCd.D4 lineage is missing, as indicated in grey; see Supplementary Fig. 5). Scale bar (for **c** and **d**), 200 μ m.

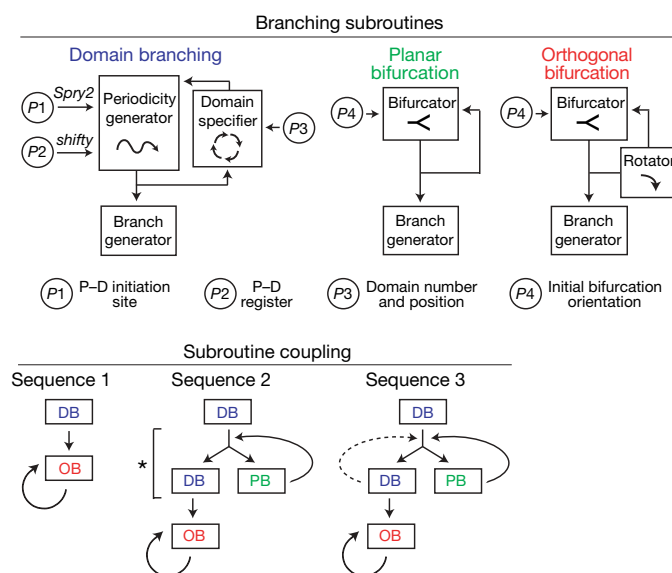


Figure 6 | A formal model of the lung branching programme. Top, representation of branching modes as subroutines using different combinations of patterning and morphogenesis operations (boxed; see Fig. 2f). Subroutines are locally modified by input parameters $P1$ – $P4$ that regulate the variables indicated. P – D , proximal–distal. Bottom, three subroutine-coupling schemes generate the three observed sequences of subroutine use (Fig. 3c). The schemes are related: bypass of the step marked by the asterisk in sequence 2 generates sequence 1, and repeat of the domain branching step (dashed line) generates sequence 3.

alveolus formation, and patterning pulmonary blood vessels. Because airways seem to set arterial and smooth muscle pattern (R. J. Metzger *et al.*, in preparation), and signals from the airways probably direct morphogenesis of other tissues in the lung too (F. H. Espinoza *et al.*, in preparation), parsing the airway branching programme is a critical step towards elucidating the full programme of lung development.

Evolution of branching networks. Branching networks come in many sizes, cellular architectures and branching complexities that differ between organs and species^{35–37}. For example, the human bronchial tree contains millions of branches, several orders of magnitude more than in mouse, whereas the lobes of frog lungs are unbranched sacs. Human and mouse lungs also differ in lobation and branch pattern.

The modular logic of the mouse lung branching programme suggests how such structural diversity is created during evolution^{38–40}. New branching patterns can arise by reiterative use of subroutines or new patterns of their deployment. Indeed, although limited, the developmental data available for human and pig provide evidence that at least domain branching and orthogonal bifurcation are used in other animals^{20,21,41}. New branching patterns can also arise by local modifications of subroutines, like the increased number and altered positions of domain branches in *Spry2* and shifty mutants, and the reduction of the standard four-domain structure to three domains for the accessory lobe branch R.Ac. More extreme modifications could create entirely new subroutines: orthogonal bifurcation may have evolved from planar bifurcation by acquisition of the rotator function. Branching subroutines controlled by a master routine may represent a general biological strategy for encoding and evolving complex branch patterns.

METHODS SUMMARY

Outbred CD1 embryos, inbred A/J, C3H/HeJ, C57BL/6J, DBA/2J and FVB/NJ embryos, and *iv* (ref. 27) and *Spry2*^{ΔORF/ΔORF} (ref. 28) null and littermate control embryos were dissected in PBS. Noon of the day a vaginal plug was detected was considered ~E0.5. Lungs were fixed and airway epithelium was visualized by indirect immunofluorescence after staining with rat anti-E-cadherin primary antibody (clone ECCD-2)⁴², biotinylated secondary antibody, and avidin-peroxidase and tyramide histochemistry. Some lungs were double stained with Cy3-conjugated mouse anti-smooth-muscle- α -actin antibody (clone 1A4) to enhance visualization of early branch generations. See Methods and Supplementary Methods for details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 December 2007; accepted 15 April 2008.

Published online 7 May 2008.

- Weibel, E. R. *The Pathway for Oxygen* (Harvard Univ. Press, Cambridge, Massachusetts, 1984).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- Bejan, A. *Shape and Structure, From Engineering to Nature* (Cambridge Univ. Press, Cambridge, 2000).
- Mauroy, B., Filoche, M., Weibel, E. R. & Sapoval, B. An optimal bronchial tree may be dangerous. *Nature* **427**, 633–636 (2004).
- Aeby, C. *Der Bronchialbaum der Säugethiere und des Menschen, nebst Bemerkungen über den Bronchialbaum der Vögel und Reptilien* (Engelmann, Leipzig, 1880).
- Boyden, E. A. *Segmental Anatomy of the Lungs; a Study of the Patterns of the Segmental Bronchi and Related Pulmonary Vessels* (McGraw-Hill, New York, 1955).
- Weibel, E. R. & Gomez, D. M. Architecture of the human lung. *Science* **137**, 577–585 (1962).
- Weibel, E. R. *Morphometry of the Human Lung* (Academic, New York, 1963).
- Metzger, R. J. & Krasnow, M. A. Genetic control of branching morphogenesis. *Science* **284**, 1635–1639 (1999).
- Meinhardt, H. Morphogenesis of lines and nets. *Differentiation* **6**, 117–123 (1976).
- Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1983).
- Nelson, T. R. & Manchester, D. K. Modeling of lung morphogenesis using fractal geometries. *IEEE Trans. Med. Imaging* **7**, 321–327 (1988).
- Prusinkiewicz, P. & Lindenmayer, A. *The Algorithmic Beauty of Plants* (Springer, New York, 1990).
- Kitaoka, H., Takaki, R. & Suki, B. A three-dimensional model of the human airway tree. *J. Appl. Physiol.* **87**, 2207–2217 (1999).

- Tawhai, M., Pullan, A. & Hunter, P. Generation of an anatomically based three-dimensional model of the conducting airways. *Ann. Biomed. Eng.* **28**, 793–802 (2000).
- Miura, T. Modeling lung branching morphogenesis. *Curr. Top. Dev. Biol.* **81**, 291–310 (2007).
- Tebockhorst, S., Lee, D., Wexler, A. S. & Oldham, M. J. Interaction of epithelium with mesenchyme affects global features of lung architecture: a computer model of development. *J. Appl. Physiol.* **102**, 294–305 (2007).
- Cardoso, W. V. & Lu, J. Regulation of early lung morphogenesis: questions, facts and controversies. *Development* **133**, 1611–1624 (2006).
- His, W. Zur Bildungsgeschichte der Lungen beim menschlichen Embryo. *Arch. Anat. Entwicklungsgeschichte* **1887**, 89–106 (1887).
- Flint, J. M. Development of the lungs. *Am. J. Anat.* **6**, 1–138 (1906).
- Heiss, R. Zur Entwicklung und Anatomie der menschlichen Lunge. *Arch. Anat. Physiol. Anatomische Abteilung* **1919**, 1–129 (1919).
- Borghese, E. The development *in vitro* of the submandibular and sublingual glands of *Mus musculus*. *J. Anat.* **84**, 287–302 (1950).
- Alescio, T. Osservazioni su culture organotipiche di polmone embrionale di topo. *Arch. Ital. Anat. Embriol.* **65**, 323–363 (1960).
- Massoud, E. A. *et al.* *In vitro* branching morphogenesis of the fetal rat lung. *Pediatr. Pulmonol.* **15**, 89–97 (1993).
- Watanabe, T. & Costantini, F. Real-time analysis of ureteric bud branching morphogenesis *in vitro*. *Dev. Biol.* **271**, 98–108 (2004).
- Supp, D. M., Witte, D. P., Potter, S. S. & Brueckner, M. Mutation of an axonemal dynein affects left–right asymmetry in *inversus viscerum* mice. *Nature* **389**, 963–966 (1997).
- Hummel, K. P. & Chapman, D. B. Visceral inversion and associated anomalies in the mouse. *J. Hered.* **50**, 9–13 (1959).
- Shim, K., Minowada, G., Coling, D. E. & Martin, G. R. *Sprouty2*, a mouse deafness gene, regulates cell fate decisions in the auditory sensory epithelium by antagonizing FGF signaling. *Dev. Cell* **8**, 553–564 (2005).
- Hacohen, N. *et al.* *Sprouty* encodes a novel antagonist of FGF signaling that patterns apical branching of the *Drosophila* airways. *Cell* **92**, 253–263 (1998).
- Minowada, G. *et al.* Vertebrate *Sprouty* genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. *Development* **126**, 4465–4475 (1999).
- Eppig, J. T. *et al.* *Mouse Genome Database* (<http://www.informatics.jax.org>) (2005).
- Hogan, B. L. Morphogenesis. *Cell* **96**, 225–233 (1999).
- Warburton, D. *et al.* Molecular mechanisms of early lung specification and branching morphogenesis. *Pediatr. Res.* **57**, 26R–37R (2005).
- Maeda, Y., Dave, V. & Whitsett, J. A. Transcriptional control of lung morphogenesis. *Physiol. Rev.* **87**, 219–244 (2007).
- al-Awqati, Q. & Goldberg, M. R. Architectural patterns in branching morphogenesis in the kidney. *Kidney Int.* **54**, 1832–1842 (1998).
- Lu, P., Sternlicht, M. D. & Werb, Z. Comparative mechanisms of branching morphogenesis in diverse systems. *J. Mammary Gland Biol. Neoplasia* **11**, 213–228 (2006).
- Thomson, A. A. & Marker, P. C. Branching morphogenesis in the prostate gland and seminal vesicles. *Differentiation* **74**, 382–392 (2006).
- Shubin, N. H. & Alberch, P. A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* **20**, 319–387 (1986).
- Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
- Prusinkiewicz, P. *et al.* Evolution and development of inflorescence architectures. *Science* **316**, 1452–1456 (2007).
- Wells, L. J. & Boyden, E. A. The development of the bronchopulmonary segments in human embryos of horizons XVII to XIX. *Am. J. Anat.* **95**, 163–201 (1954).
- Hirai, Y., Nose, A., Kobayashi, S. & Takeichi, M. Expression and role of E- and P-cadherin adhesion molecules in embryonic histogenesis. I. Lung epithelial morphogenesis. *Development* **105**, 263–270 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Krasnow laboratory and P. Brown, D. Brutlag, N. Hacohen, D. Kingsley, L. Mündermann, J. Spudich and E. Storm for advice and discussion, and M. Kumar and M. Petersen for help with preparing the figures. This work was funded by grants from National Institutes of Health (to M.A.K. and G.R.M.). M.A.K. is an investigator of the Howard Hughes Medical Institute.

Author Contributions R.J.M. and M.A.K. conceived the experiments. R.J.M. designed and performed experiments and collected data. O.D.K. and G.R.M. contributed to conception and design of the *Spry2* experiments and provided genotyped *Spry2* embryos. R.J.M. and M.A.K. analysed the data and wrote the manuscript. All authors discussed results and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.J.M. (ross.metzger@ucsf.edu) or M.A.K. (krasnow@cmgm.stanford.edu).

METHODS

Mice. CD1 mice (Charles River Laboratories) were used for the wild-type analysis. Inbred strains (Jackson Laboratory) were: A/J ($n = 21$ lungs analysed), C3H/HeJ ($n = 8$), C57BL/6J ($n = 20$), DBA/2J ($n = 10$) and FVB/NJ ($n = 13$) mice. *iv/iv* mice²⁷ were from the Jackson Laboratory. *Spry2*^{-/-} embryos and their *Spry2*^{+/-} and *Spry2*^{-/-} littermates were obtained from crosses of *Spry2* mutant alleles²⁸ on three different mixed genetic backgrounds; similar results were obtained for all three.

Imaging. Specimens were imaged on Leica MZFLIII, Leica MZ16FA or M²Bio (Kramer Scientific) fluorescence stereomicroscopes. Images were captured with a Spot RT slider (Diagnostic Instruments) camera with Spot software or a Retiga 2000R (Q Imaging) camera with Image-Pro (Media Cybernetics) software. Adobe Photoshop software was used to adjust image levels and to pseudocolour the images in Fig. 3a.

Constructing the branch lineage of the mouse bronchial tree. For each branch, the lineage was reconstructed by assembling the branch patterns of fixed, immunostained lungs from E11 up to E15 (as described in the text) into local, self-consistent orders from which we could infer the dynamic sequence and pattern of branching. For most branches, the lineage was based on groups of specimens that included all or most intermediates. However, for some branches that form during E14, we did not obtain every intermediate but were nevertheless able to reconstruct the lineage and branching dynamics based on the morphological similarity of the intermediates obtained to those of positions where we were able to reconstruct the lineage in detail. Lineages L.M2–3, RCd.M2, R.Ac.A1–4 and R.RMid.V1–4 were the most difficult to reconstruct. They form relatively late and, because of the shape of the lobes, become difficult to visualize clearly as branching proceeds.

Representing the branch lineage in the lineage diagram. Because there is variability among specimens in the local rate of progression through the lineage (see text and Supplementary Fig. 2), and because branching in most positions continues beyond E15, the extent of the lineage shown in Supplementary Fig. 1 was, with two exceptions, based on a single, representative E15 lung. One exception was the number of branches in each domain off L and RCd. The number of secondary branches in these eight domains appears to be complete by E14 because the number of branches did not increase between E14 and E15 as it did for other domains. Hence, for these eight domains, we show in the lineage diagram the maximum number of secondary branches observed in five different E15 lungs. The other exception was the number of rounds of orthogonal bifurcation shown for each sublineage. In the lineage diagram, we show the same number of rounds of orthogonal bifurcation for each daughter branch even though sister branches do not always bifurcate synchronously, and we show a maximum of four rounds because it was difficult to ascertain the branching pattern beyond that. The R.Md.D1 'optional' lineage (see text) was present in the E15 specimen used and hence is included in the lineage diagram, but other optional lineages were not, including R.Ac.P1, and anterior or posterior tertiary branches off L.L2–6 and RCd.L1–5. Branching 'errors' (see text and Fig. 4b–e) are also not included in the diagram.

Because there is some variability in the absolute orientation of orthogonal bifurcations (see text), particularly late in a series of orthogonal bifurcations, in the lineage diagram we did not name the branches produced by the fourth round of orthogonal bifurcations by their orientation, as we did for the earlier rounds, but instead used an asterisk. For the earlier rounds, the orientation given is the one most frequently observed.

Assigning branching modes. Assigning the mode of formation to each branch in the lineage was obvious for most branches from the branching dynamics inferred from the series of developmental intermediates. However, the following branches warrant special comment.

Some branches that appear to form by domain branching (for example, RCd.V1 and L.V1; see text) were 'singletons'; that is, they were the only branch found in a particular domain. We assume that these are unusual domains in which just a single branch typically forms or are domains in which additional branches form later in development, hence we assigned these branches as the first branch in the domain (for example, L.L2.A.D1). Likewise, some entire domains that are missing in the lineage presumably form later in development. However, there may also be more variation in the number and position of domains that form late.

Because bifurcations were identified as being planar or orthogonal based on the orientation of the subsequent round of bifurcation, for the final generation in lineages in which we did not collect information on subsequent branching events, we designated the bifurcation as orthogonal or planar based on morphology and analogy to neighbouring positions in the lineage. For example, L.L2.A.A.V1.A and L.L2.A.A.V1.P, which form off a branch that forms by domain branching, look like the first round of orthogonal bifurcation as seen off other branches that form by domain branching.

In several positions in the lineage there were potential ambiguities in assigning branching modes. One was at primary branch tips, where several branches we assigned as forming by domain branching (RCd.L4, RCd.L5, L.5 and L.6) could also be interpreted as planar bifurcations based on branching dynamics and morphology. However, this alternative interpretation is unlikely because elsewhere in the developing lung branches that form by planar bifurcation undergo domain branching in the same domains used by the parent, whereas these branches use only a subset of the domains used by the parent branches. It is also possible based on branching morphology to interpret RCd.L5 and L.L6 as simply the continuation of the primary branch. However, this too is unlikely because they only form daughter branches in two of the four domains of the primary branch.

In lineages using sequence 3 (for example, L.L1), where a domain (for example, L.L1.A1 and L.L1.A2) and daughter branches that form by planar bifurcation (for example, L.L1.A and L.L1.P) lie in the same plane, domain branches could alternatively be interpreted as forming by highly asymmetric planar bifurcations in which one daughter branch of the bifurcation appears to form part of the domain and the other appears to be a continuation of the parent branch. However, we do not prefer this interpretation because all other planar bifurcations are symmetric. Rather, we suggest that this relatedness indicates how domain branching may have evolved from planar bifurcation.

A two-tiered mechanism for stabilization and immobilization of E-cadherin

Matthieu Cavey¹, Matteo Rauzi^{1,2}, Pierre-François Lenne² & Thomas Lecuit¹

Epithelial tissues maintain a robust architecture which is important for their barrier function, but they are also remodelled through the reorganization of cell–cell contacts. Tissue stability requires intercellular adhesion mediated by E-cadherin, in particular its trans-association in homophilic complexes supported by actin filaments through β - and α -catenin. How α -catenin dynamic interactions between E-cadherin/ β -catenin and cortical actin control both stability and remodelling of adhesion is unclear. Here we focus on *Drosophila* homophilic E-cadherin complexes rather than total E-cadherin, including diffusing ‘free’ E-cadherin, because these complexes are a better proxy for adhesion. We find that E-cadherin complexes partition in very stable microdomains (that is, bona fide adhesive foci which are more stable than remodelling contacts). Furthermore, we find that stability and mobility of these microdomains depend on two actin populations: small, stable actin patches concentrate at homophilic E-cadherin clusters, whereas a rapidly turning over, contractile network constrains their lateral movement by a tethering mechanism. α -Catenin controls epithelial architecture mainly through regulation of the mobility of homophilic clusters and it is largely dispensable for their stability. Uncoupling stability and mobility of E-cadherin complexes suggests that stable epithelia may remodel through the regulated mobility of very stable adhesive foci.

Intercellular adhesion mediated by cadherin molecules such as E-cadherin (E-cad) has a central role in the maintenance of epithelial polarity^{1,2} and tissue architecture^{3,4}. Loss of E-cad leads to a collapse of epithelial tissues and their transformation to multilayered mesenchymes. However, epithelia can remodel extensively during development and in a range of organisms. How adhesion between cells controls tight cell association and allows contact remodelling is unknown. It is believed that this dual property resides in the molecular dynamics of E-cad-based adhesion.

E-cad exists in at least two main pools at the surface of epithelial cells: a monomer pool that exchanges with a trans-homophilic dimer pool at contacting surfaces⁵, hereafter called homo-E-cad. Adhesion strengthening between cells is believed to require homo-E-cad clustering and stabilization by actin filaments (F-actin) and β -catenin (β -Cat)⁶. How homo-E-cad is distributed and dynamically remodelled in living epithelia is unclear.

Biochemical and cell culture studies in the past few decades have led to a largely accepted framework whereby homo-E-cad clustering requires β -Cat and α -catenin (α -Cat) at its core, as well as actin filaments, which limit the diffusion of and may also stabilize homo-E-cad⁶. Single-particle tracking of E-cad molecules in non-epithelial cells has shown that E-cad diffusion is indeed limited by its ability to interact with F-actin⁷. This simple framework was challenged recently when it was shown in cultured epithelial cells that diffusion of E-cad fused to green fluorescent protein (E-cad–GFP) was independent of F-actin, and that α -Cat bound to either F-actin or E-cad– β -Cat complexes but not both simultaneously, allowing for a dynamic linkage between E-cad and F-actin⁸. Although this offered a molecular basis for the dynamics of adhesion, it became difficult to understand in this context how adhesion may be strengthened by F-actin. Until now, relevant E-cad dynamics using E-cad–GFP fusion proteins were inferred from the average effect of monomeric and homo-E-cad, without distinction of either, in large (few micrometres) areas of the zonula adherens^{8,9} at various stages of junction formation¹⁰. Here we extend

these studies and resolve the contribution of the two populations in a fully functional living epithelial layer of the *Drosophila* embryo that undergoes extensive remodelling. In particular, we focused our analysis on homo-E-cad because this corresponds to the E-cad population most directly responsible for adhesion strength at any given time.

Homophilic E-cadherin clusters in very stable microdomains

As initially reported, in early embryonic epithelia, E-cad (encoded by the *shg* gene in *Drosophila*), β -Cat (called Armadillo (Arm) in *Drosophila*^{11–14}) and α -Cat form brightly fluorescent spots (Fig. 1a–c). Similarly sized electron-dense microdomains called spot adherens junctions (SAJs)¹⁵ exhibit more closely apposed membranes, suggesting that E-cad spots may be sites of homo-E-cad enrichment. E-cad clusters increased in density but persisted throughout embryogenesis (Fig. 1d) and at all stages examined; homo-E-cad was neither randomly nor uniformly distributed. We characterized the dynamics of E-cad within and outside SAJs using a functional E-cad–GFP fusion protein¹⁶. E-cad–GFP formed fluorescent spots in the zonula adherens which were indistinguishable from the endogenous E-cad spots (Fig. 1e). We analysed its molecular dynamics using fluorescent recovery after photobleaching (FRAP). By performing FRAP in very small areas (840-nm diameter) we found that E-cad–GFP dynamics within and outside SAJs were different. Representative kymographs (showing the temporal evolution of a fluorescent intensity profile along a cell junction; Fig. 1f, g) and signal recovery curves (Fig. 1h) are shown for a single E-cad–GFP spot (Fig. 1f, h) and a low fluorescent E-cad–GFP area (Fig. 1g, h). The diffusion coefficient ranged from 0.5×10^{-2} to $6 \times 10^{-2} \mu\text{m}^2 \text{s}^{-1}$, consistent with published data^{7,17}, and was unchanged when actin was depolymerized with latrunculin-A (Lat-A; not shown), consistent with previous reports⁸. Notably, full fluorescence of bright spots never recovered within 60 s of FRAP (Fig. 1f, h) and the mobile fraction was lower in bright spots (<45%) than in the surrounding areas (up to 80%, Fig. 1i). To characterize turnover rates and the

¹Institut de Biologie du Développement de Marseille Luminy, UMR 6216 CNRS-Université de la Méditerranée, Campus de Luminy case 907, 13288 Marseille Cedex 09, France. ²Institut Fresnel, UMR 6133 CNRS-Université Paul Cézanne Aix-Marseille III, Domaine Universitaire de Saint Jérôme, 13397 Marseille Cedex 20, France.

stability of SAJs over longer time courses, we made a fusion between E-cad and the photoconvertible protein EosFP¹⁸. E-cad–EosFP was distributed in spots (Fig. 1j) similar to endogenous E-cad that form within a few minutes from a pre-existing pool (Supplementary Fig. 1). We chased the stability of homo-E-cad and found that once formed, photo-converted homo-E-cad persisted in SAJs for about 1 h (Fig. 1j). Quantification showed no detectable exchanges between photo-converted SAJs and their surroundings (Supplementary Fig. 2). Thus, E-cad is indeed highly dynamic within the zonula adherens (diffusing monomer), but clusters of homo-E-cad in SAJs form very stable complexes, consistent with the idea that SAJs are bona fide sites of adhesion. This realization prompted a focus on SAJ stability and mobility.

Two dynamically distinct actin populations

A mosaic distribution of homo-E-cad is not predicted at thermodynamic equilibrium as the line tension at the perimeter of SAJs with the surrounding membrane should favour larger domains. Active mechanisms must thus be at work to maintain SAJs. This led us to investigate the function of actin in adhesion. As actin limits diffusion of single E-cad molecules due to its ability to interact with E-cad– β -Cat complexes through α -Cat (a fact supported by, for example,

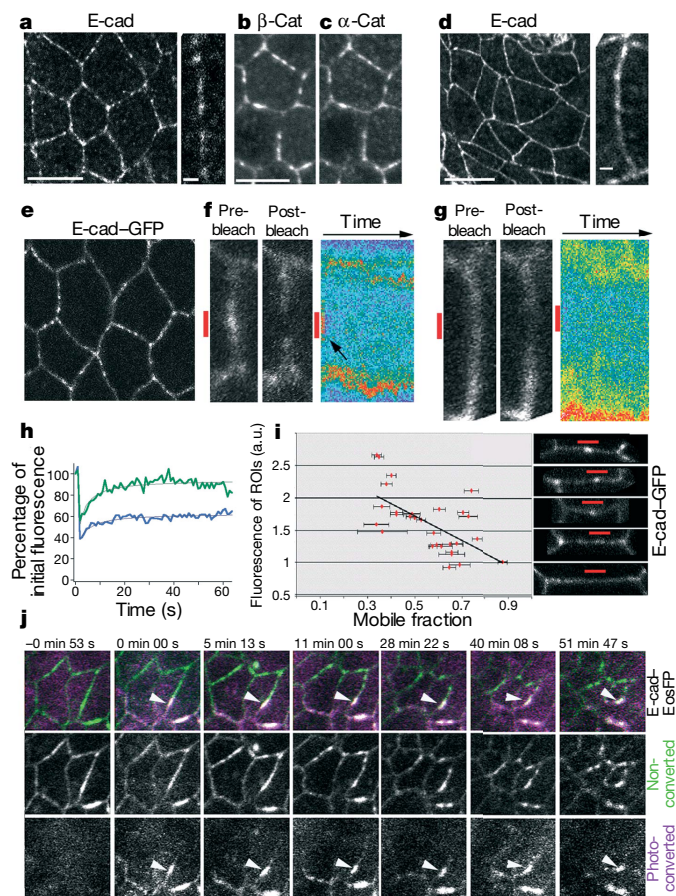


Figure 1 | Homophilic E-cad clusters in highly stable microdomains. **a–g**, Spots of E-cad, β -Cat, α -Cat and E-cad–GFP are visible at all stages of embryogenesis (**a–c**, **e–g**, stage 8; **d**, stage 14). Scale bars: 5 μ m (**a** (left panel only), **b**, **c**, **d** (left panel only), **e**); 1 μ m (right panel only of **a** and **d**). **f–i**, FRAPs of E-cad–GFP showing the absence of recovery for bright (**f**) but not low-intensity (**g**) regions. Right panels are corresponding kymographs showing the evolution over 90 s of fluorescence intensity along the junctions. Red bars denote region of bleach. **h**, Fluorescence recovery curves for the examples shown in **f** (blue) and **g** (green). **i**, Mobile fractions versus pre-bleach brightness of the region of interest (ROI); representative examples shown in the right panel). **j**, Photoconverted spots of E-cad–EosFP (white arrowheads; stage 9–10) do not exchange with neighbouring regions for >50 min (see Supplementary Fig. 2).

single-particle tracking in non-epithelial cells), the traditional view states that actin thereby stabilizes E-cad homophilic complexes. High-resolution confocal imaging of F-actin using phalloidin shows a dense F-actin network at the zonula adherens (Fig. 2a, b).

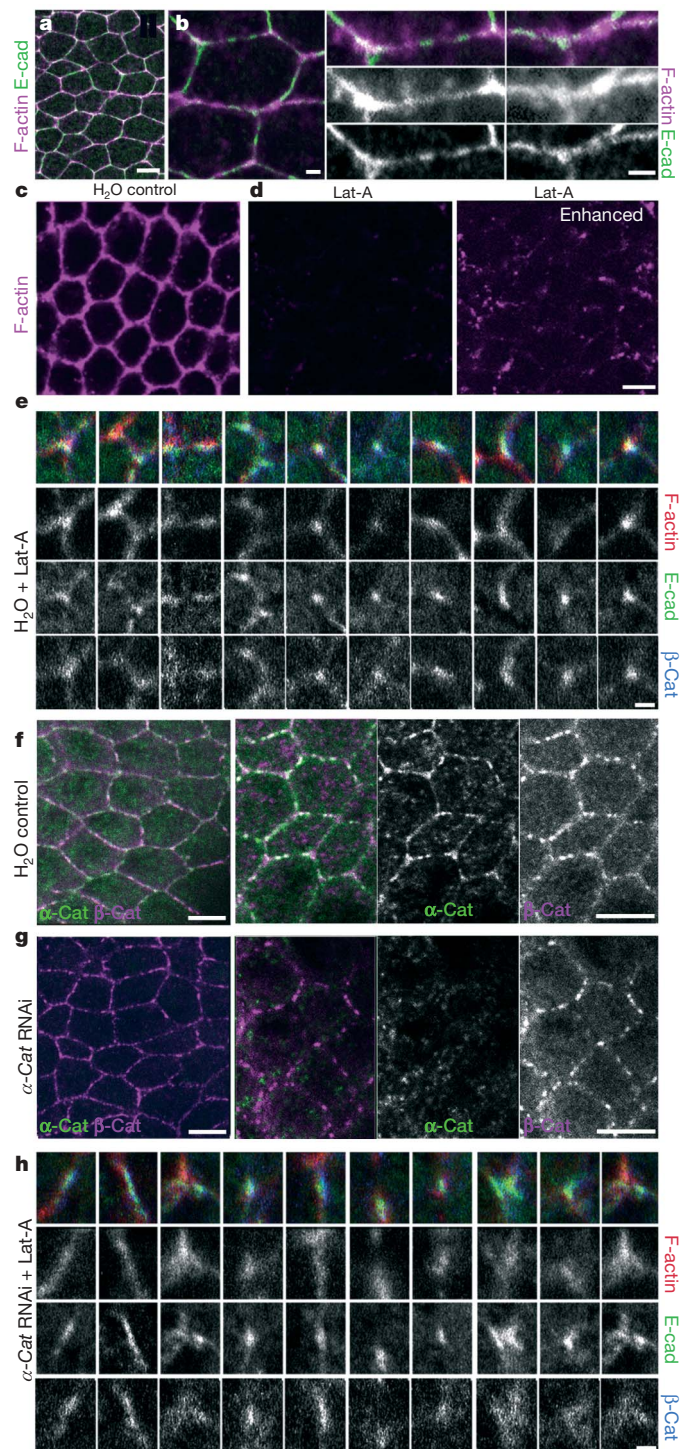


Figure 2 | A stable pool of actin localizes at and stabilizes SAJs independently of α -Cat. **a**, **b**, F-actin and E-cad intensities are not correlated. **c**, **d**, Latrunculin-A (Lat-A) injection depletes most of F-actin (**d**, signal enhanced in right image) compared to control H₂O-injected embryos (**c**). **e**, Lat-A-resistant F-actin patches co-localize with remaining E-cad and β -cat spots. **f**, **g**, α -Cat protein is strongly depleted by α -Cat RNAi injection, whereas β -cat spots are left intact. **h**, Double injection of α -Cat RNAi and subsequently Lat-A does not remove Lat-A-resistant F-actin patches from the remaining E-cad and β -Cat spots, similarly to the control H₂O and Lat-A injections (**e**). Scale bars: 5 μ m (**a**, **c**, **d**, **f**, **g**); 1 μ m (**b**, **e**, **h**).

Fluorescent quantification of F-actin showed that only a fraction (27%) of junctional actin localized within SAJs, such that F-actin is not enriched within these areas (Fig. 2b). As this only relies on steady-state distribution, we investigated the stability of F-actin using Lat-A (20 μM), which blocks polymerization of F-actin. As little as 5 min after Lat-A injection, most (>80%) F-actin was removed from cell junctions, indicating a rapid depolymerization and hence turnover rate of this actin network in the zonula adherens (Fig. 2c, d). SAJs became slightly dimmer under these conditions, but notably never disappeared and persisted for at least 45 min (Supplementary Movie 2). A small fraction of stable actin persisted (Fig. 2d) and localized at

SAJs together with E-cad and β -Cat (Fig. 2e). We conclude that two populations of F-actin, with different dynamic properties, coexist at the zonula adherens: small and stable actin patches within SAJs, and a predominant network that constantly recycles at the zonula adherens. The latter pool is not required for homo-E-cad stability at SAJs. Owing to our inability to depolymerize actin fully at SAJs with either Lat-A or cytochalasin-D (not shown), we cannot firmly demonstrate that actin patches are essential for homo-E-cad stability. However, we think actin patches underlie SAJs stability, first because actin patches and SAJs are both very stable, and second because SAJs are slightly dimmer in Lat-A-treated embryos compared to controls. Note that our experiments only address the stability of SAJs once they are formed. Their formation is very rapid (Supplementary Fig. 1) and may depend on mechanisms other than stability.

α -Cat is not essential for homo-E-cad stability

α -Cat is believed to mediate, albeit in a dynamic way⁸, E-cad– β -Cat and F-actin interactions; however, earlier studies had not fully resolved its precise role in adhesion^{19–21}. The knockdown of α -Cat using RNA interference (RNAi) removed all detectable α -Cat in SAJs on the basis of immunofluorescence (Fig. 2f, g) and at least 93% total α -Cat in embryos on the basis of western blotting (Supplementary Fig. 3). After this knockdown, the epithelial monolayer collapsed markedly (not shown). However, in α -Cat RNAi embryos bright E-cad– β -Cat spots were detected as in controls and could be traced for over 45 min during tissue collapse (Figs 2g and 3; see also Supplementary Movies 1 and 3), suggesting that the primary function of α -Cat in adhesion is not to control homo-E-cad stability in SAJs. Consistent with this, α -Cat was not even required to maintain the Lat-A-resistant F-actin patches at SAJs (Fig. 2h). Although we cannot totally rule out that undetectable α -Cat at SAJs may contribute to their stability to a small extent, our data clearly demonstrate that other molecules must underlie the strong association of stable actin patches to SAJs and homo-E-cad stability in SAJs. Furthermore, the marked role of α -Cat in epithelial stability reflects a different molecular function than control of SAJ stability. These results led us to shift our attention to the dynamic actin network that underlies the zonula adherens and to investigate the function of α -Cat in this context.

A tensile actin network controls mobility of SAJs

Although SAJs were still present in Lat-A-treated embryos, the epithelium collapsed rapidly, forming a multilayered mesenchyme, suggesting that defects occur in the dynamic distribution of SAJs in the zonula adherens. Similar defects occurred when the junctional actin network was fragmented in mutants for *bitesize* (*btsz*), a gene encoding a synaptotagmin-like protein previously shown to organize actin filaments at cell junctions through the F-actin-binding protein Moesin²². We quantified the movement of individual SAJs within the zonula adherens. Kymograph analysis of E-cad–GFP spots during gastrulation consistently showed similar trajectories (Fig. 3a). We measured the distance d_i between spots (inter-spot distance) as a function of time in the zonula adherens (Fig. 3b), circumventing a bias introduced by the lateral drift of the epithelial tissue. The mean square displacement (MSD = $\langle d_i^2 \rangle$) of d_i showed a linear relationship as a function of time (not shown), yielding the diffusion coefficient (D) of SAJs^{23,24} (Fig. 3c, d). 64.3% of spots were classified as stationary ($n = 84$) as their apparent diffusion coefficients were below the detection limit ($D < 1.6 \times 10^{-3} \mu\text{m}^2 \text{s}^{-1}$), which we determined by similarly analysing paraformaldehyde-fixed immobile homo-E-cad spots (Fig. 3d, e, red). The remaining mobile SAJs (35.7%) exhibited very low diffusion coefficients ($D < 6.3 \times 10^{-3} \mu\text{m}^2 \text{s}^{-1}$), indicating that SAJs have a constrained motion at the junctional cortex. In both *btsz* RNAi and Lat-A-injected embryos, individual SAJs visible in kymographs (Fig. 3a) were moving faster within the plane of the zonula adherens. 65% ($n = 74$) became mobile in *btsz* mutants and over 83% ($n = 119$) in Lat-A-injected

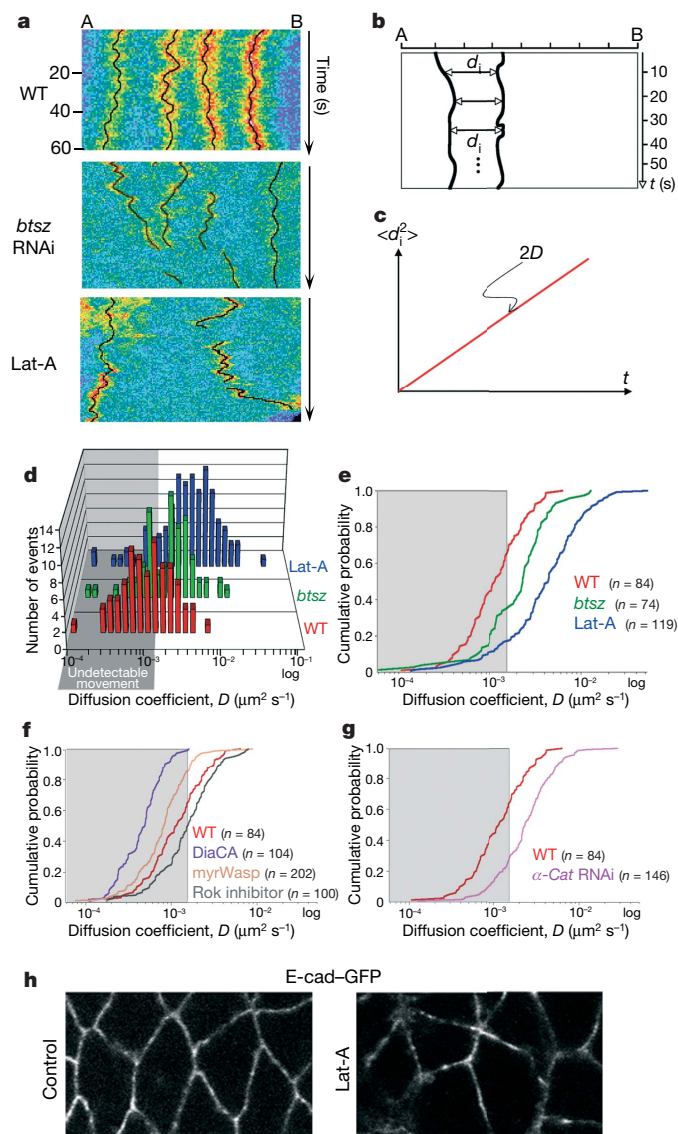


Figure 3 | Lateral mobility of SAJs controlled by a dynamic actin network. **a**, Trajectories of SAJs on representative kymographs of wild type (WT), *btsz* RNAi and Lat-A-injected embryos; time is indicated in seconds. **b**, **c**, The distance between two spots is measured at different time intervals (**b**) to extract the mean square displacement (MSD; $\langle d_i^2 \rangle$), which is linearly proportional to the time for diffusive motion (**c**) and gives the diffusion coefficient, D . **d**, **e**, Histograms (**d**) and cumulative probability distributions (**e**) of D for wild type, *btsz* and Lat-A embryos (log scale). Shaded areas indicate that D is below the detection limit ($1.6 \times 10^{-3} \mu\text{m}^2 \text{s}^{-1}$). **f**, **g**, Excess polymerization of actin by Diaphanous (DiaCA) and Wasp (myrWasp) (**f**) further immobilizes E-cad–GFP spots compared to wild type; in contrast, depletion of α -Cat (**g**) or inhibition of myosin-II contractility using a Rok inhibitor (**f**) results in enhanced mobility. **h**, Loss of uniform distribution of SAJs in Lat-A-injected embryos.

embryos (Fig. 3d, e and Supplementary Movie 2). Correspondingly, SAJs had higher diffusion coefficients than controls ($P = 1.7 \times 10^{-5}$ for *btsz* RNAi and $P = 3.6 \times 10^{-16}$ for Lat-A). We also measured shorter residence time of SAJs in 1- μm -thick confocal sections when actin was disrupted (*btsz* mutant and Lat-A injection) compared to wild-type embryos (not shown), demonstrating that SAJs also diffuse away from junctions along the apical-basal axis in these conditions. As a result, SAJs lost their uniform distribution around the cell periphery (Fig. 3h) and cells failed to maintain proper uniform adhesion, leading to epithelium collapse. Conversely, forcing the polymerization of the actin network through expression of a constitutively active form of Diaphanous (also known as Formin)²⁵ to promote unbranched actin elongation²⁶, or expression of a myristoylated form of Wasp to promote cortical actin branching²⁷, caused significant ($P = 9.57 \times 10^{-13}$ and $P = 4.87 \times 10^{-4}$, respectively) further

immobilization of SAJs compared to wild type (Fig. 3f). In addition, removing myosin-II motors within the cortical actin network using a Rho-associated kinase (Rok) inhibitor increased SAJ mobility (Fig. 3f, $P = 3 \times 10^{-2}$). A simple interpretation of this is that actin network stiffening by myosin-II motor and crosslinking activity is required for SAJ immobilization. Taking these results together, we conclude that the dynamic actin network at the zonula adherens controls the lateral mobility of SAJs but not homo-E-cad clustering and molecular stability.

α -Cat mediates tethering of homo-E-cad by cortical actin

Two mechanisms could underlie the constrained mobility of SAJs. The dynamic actin network could prevent lateral diffusion of SAJs by forming non-specific obstacles as in a corral. Conversely, SAJs may be tethered to the actin network by specific interactions. Our findings support the latter mechanism and are inconsistent with the former mechanism playing a major part. First, we show that α -Cat serves as a specific dynamic tether between SAJs and actin. Although the stability of SAJs does not require α -Cat, their individual dynamics increased significantly in α -Cat RNAi embryos (Fig. 3g and Supplementary Movie 3; $P = 3.87 \times 10^{-8}$), explaining the tissue collapse as also observed in Lat-A-injected embryos. To probe further the mechanical links between SAJs and the underlying actin network, we developed 'nano-scissors' to sever this network and followed the redistribution of SAJs. A near-infrared pulsed laser delivered nano-joule pulses to ablate within sub-diffraction-limited volumes at the cortex²⁸. Short (~ 5 ms) ablation led to the progressive (within 60 s) redistribution, away from the point of ablation, of the junctional actin network visualized by fusion of GFP to the actin-binding domain of moesin (MoeABD-GFP²⁹; Fig. 4a). Three-photon uncaging of caged fluorescein ensured that membranes were still impermeable and that actin redistribution was not caused by a hole in the plasma membrane (Fig. 4a, second image from right). In such conditions, nano-ablation caused a complete redistribution of E-cad-GFP away from the point of ablation (Fig. 4b) within intact junctional cell contacts (Fig. 4b, second image from right). The redistribution of SAJs did not occur when actin was depolymerized with Lat-A (Fig. 4d) and required the myosin-II-dependent contractility of the cortical actin network (Fig. 4e), suggesting that E-cad redistribution after ablation is caused by the reorganization of actin. Moreover, when α -Cat was knocked down, SAJs did not leave the junctions where ablation occurred (Fig. 4f), consistent with α -Cat being a specific mediator of SAJ-actin interactions. A control transmembrane protein tagged with GFP (VSVG-GFP), which does not bind actin, never redistributed (Fig. 4c), further confirming that E-cad redistribution is a specific effect of actin-SAJs interactions. These results establish the primary tether function of the dynamic actin network on SAJs in the zonula adherens.

Discussion

We have identified two separable functional modules involving differently F-actin and α -Cat in stability versus mobility of homo-E-cad. Homo-E-cad is neither randomly nor uniformly distributed and clusters in very stable microdomains that persist at a variable density, at all stages of *Drosophila* embryogenesis examined, and in imaginal discs (not shown). This may reflect a general tendency towards 'mosaic' homo-E-cad organization, as observed in mammalian cells^{10,30}. We have uncovered two separable actin populations with different distributions, dynamics and functions in adhesion (Fig. 5). A small fraction of F-actin localizes at SAJs, is very stable and probably underlies stabilization of homo-E-cad. The remaining main F-actin fraction forms a very dynamic, contractile network around the junctional cortex required to restrain the mobility of SAJs within the zonula adherens by a tether mechanism. In this context, we shed light on the function of α -Cat and we find that it is dispensable for the stability of SAJs and actin stabilization at SAJs, but essential to restrict the lateral mobility of SAJs by the actin network. Failure to restrain

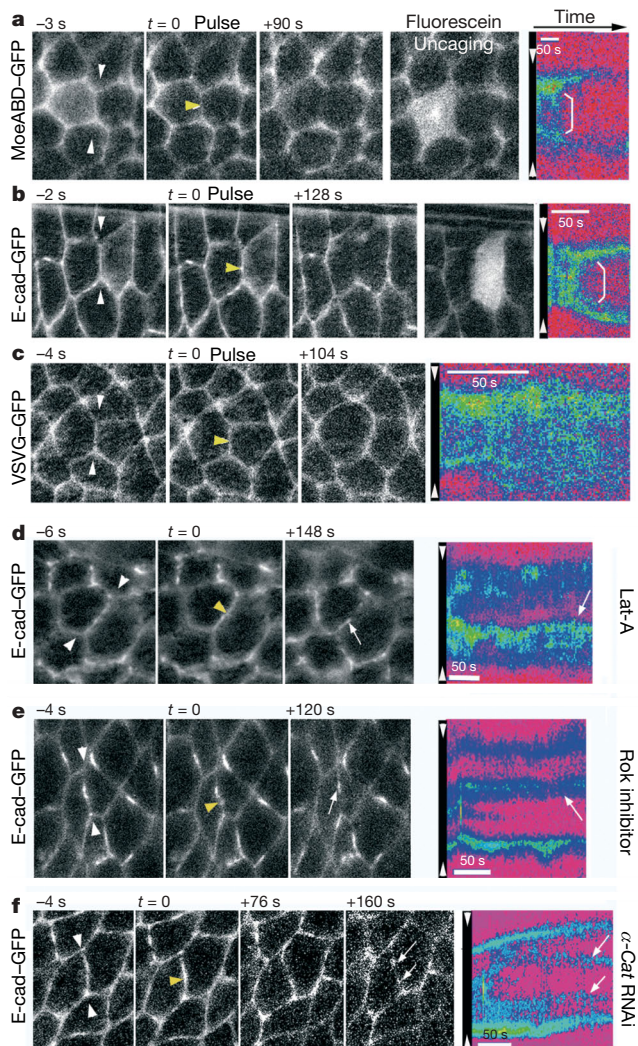


Figure 4 | Nanoablation of cortical actin network and tethering of SAJs to actin. **a–c**, Redistribution away from focal ablation spots (pulse at $t = 0$, yellow arrowheads) of F-actin (MoeABD-GFP) and E-cad-GFP but not of the transmembrane protein VSVG-GFP. **a, b**, The second panel from the right of **a** and **b** shows 3-photon uncaging of fluorescein in a single cell at the end of the experiment, demonstrating that the plasma membrane is not permeabilized by ablation. The far-right panels of **a–c** are kymographs drawn between the two white arrowheads in the corresponding far-left panels; brackets indicate clearing. **d, e**, Redistribution of E-cad-GFP requires an intact, contractile cortical actin network as it does not occur in Lat-A-injected embryos (**d**; arrows) or when myosin-II is inactivated by injection of Rok inhibitor (**e**; arrows). Redistribution is dependent on α -Cat, as indicated by the α -Cat knockdown (**f**; arrows). Time is indicated in seconds.

mobility of SAJs in α -Cat RNAi or Lat-A-treated embryos led to a collapse of the epithelium, a consequence often attributed to the destabilization of homo-E-cad complexes. These findings suggest the existence of other molecules mediating the stable association of homo-E-cad to actin patches, independent of α -Cat. Moreover, our data demonstrate that the junctional actin network pulls on SAJs via α -Cat, suggesting that α -Cat either interacts transiently with both β -Cat and actin, or that it links actin via other proteins³¹. Such individual low-affinity interactions may be impossible to detect, but their collective effects in SAJs where hundreds of E-cad molecules are clustered may be very effective.

This model has a number of important implications for the regulation of adhesion. First, stability and mobility of homo-E-cad can be mechanistically uncoupled, although both are supported by actin. It will be important to decipher how actin dynamics is differentially controlled at SAJs and at the junctional cortex. Stabilization of filaments may be differentially controlled, for example, by capping proteins, and actin patches may represent a locally stabilized pool of filaments embedded in the dynamic cortical network. Second, in light of the greater stability of SAJs (>1 h) compared to that of remodelling junctions (10–15 min)^{32,33}, contact remodelling may simply involve the regulated movement of stable SAJs towards vertices. Vertices form a topological barrier to the lateral diffusion of homo-E-cad clusters in SAJs, such that SAJs may accumulate at

vertices where they could be remodelled: this would prevent cell delamination and enable junction remodelling. Endocytosis of homo-E-cad could be another mechanism to remodel adhesion efficiently³³, although there is no consensus yet as to whether endocytosis principally targets homo-E-cad or diffusing E-cad^{34–36}. The two-tiered regulation of homo-E-cad stability and mobility thus provides a simple and coherent framework to reconcile adhesion remodelling and stability during morphogenesis.

METHODS SUMMARY

Embryo fixation, staining and injection of drugs and RNAi probes were performed as described previously²². The *btsz* probe targets nucleotides 3,065–3,556 of *btsz2* transcripts; the α -Cat probe targets nucleotides 101–828 of α -Cat transcripts.

Imaging. A Zeiss LSM 510 confocal microscope was used for fixed and live imaging as described³⁷ with the exception of nano-ablation, where classical epifluorescence with a cooled CCD camera was used.

Homophilic E-cad clusters dynamics. FRAPs of E-cad–GFP were performed with an Argon 488-nm laser over a $0.554\ \mu\text{m}^2$ area and signal recovery was tracked for 1 min. The mobile fraction was calculated by fitting signal recovery curves to the formula described previously³⁸. E-cad–EosFP photoconversion was performed with an ultraviolet (405 nm) laser at low power over regions spanning half a cell. Turnover of homo-E-cad in SAJs was then inferred by measuring the variations of non-photoconverted and photoconverted signal intensities in a cluster over a period of approximately 1 h.

Homophilic E-cad cluster mobility. The mean square displacement (MSD) of E-cad–GFP clusters was extracted from trajectories along cell contacts (during 1 to 1.5 min) using the formula described previously³⁹. We verified that their motion was diffusive using the relative deviation parameter as described²³. The diffusion coefficient, D , was then calculated assuming one-dimensional movement with the equation $\text{MSD} = 2D\Delta t$ (where Δt is the time lag over which MSD is measured). Kolmogorov–Smirnov tests were performed to determine whether the distribution of D from one sample (for example, wild type) was different from another sample (for example, Lat-A treatment).

Nano-ablation of cortical actin network. A femtosecond laser at 1,030 nm (t-pulse, Amplitude Systems) was used to ablate the cortical actin network locally. Before ablation, caged fluorescein (0.9 kDa) was injected into embryos. After ablation, uncaging of the fluorescein in one of the two cells contacting at the site of ablation ensured that the plasma membrane had not been destroyed in the process.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 November 2007; accepted 31 March 2008.

Published online 14 May 2008.

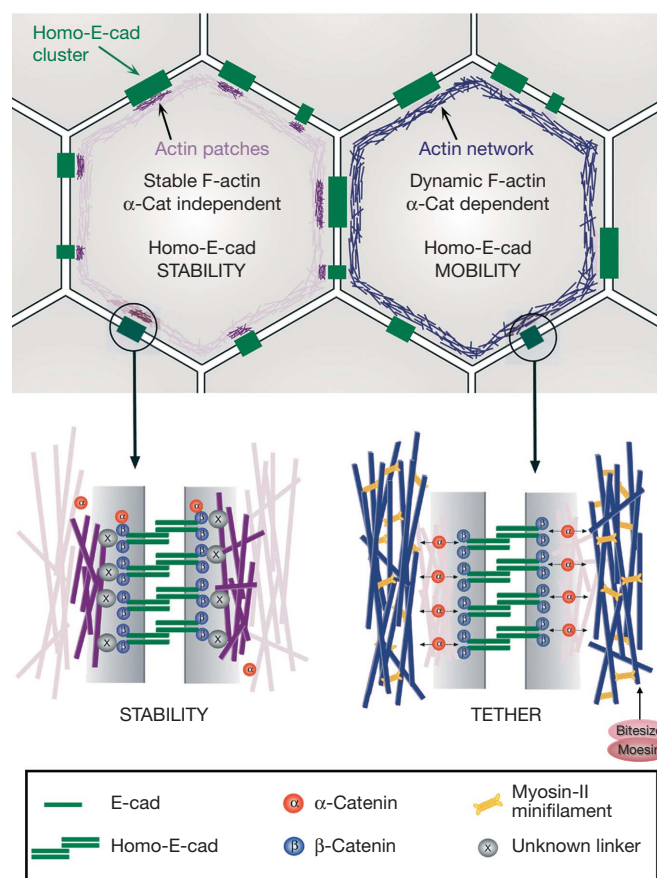


Figure 5 | Model of two-tiered regulation of homo-E-cad stability and mobility. Two actin pools regulate the stability and mobility of homo-E-cad clusters in SAJs. Stable, small actin patches concentrate and probably stabilize homo-E-cad in SAJs (left). α -Cat is not necessary for stability, and unknown linkers (X) between actin and E-cad should be involved. A dynamic, contractile network regulated by Bitesize and Moesin tethers all SAJs and limits their lateral mobility through α -Cat (right), thereby maintaining adhesion in a defined domain. Note that the two actin pools are distinct with respect to their dynamics and function and are probably intermingled at SAJs.

- Knust, E. & Bossinger, O. Composition and formation of intercellular junctions in epithelial cells. *Science* **298**, 1955–1959 (2002).
- Nelson, W. J. Adaptation of core mechanisms to generate cell polarity. *Nature* **422**, 766–774 (2003).
- Thiery, J. P. Cell adhesion in development: a complex signalling network. *Curr. Opin. Genet. Dev.* **13**, 365–371 (2003).
- Lecuit, T. & Lenne, P. F. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. *Nature Rev. Mol. Cell Biol.* **8**, 633–644 (2007).
- Leckband, D. & Prakasham, A. Mechanism and dynamics of cadherin adhesion. *Annu. Rev. Biomed. Eng.* **8**, 259–287 (2006).
- Gates, J. & Peifer, M. Can 1000 reviews be wrong? Actin, α -Catenin, and adherens junctions. *Cell* **123**, 769–772 (2005).
- Sako, Y., Nagafuchi, A., Tsukita, S., Takeichi, M. & Kusumi, A. Cytoplasmic regulation of the movement of E-cadherin on the free cell surface as studied by optical tweezers and single particle tracking: corralling and tethering by the membrane skeleton. *J. Cell Biol.* **140**, 1227–1240 (1998).
- Yamada, S., Pokutta, S., Drees, F., Weis, W. I. & Nelson, W. J. Deconstructing the cadherin-catenin-actin complex. *Cell* **123**, 889–901 (2005).
- Cliffe, A., Mieszczynek, J. & Bienz, M. Intracellular shuttling of a *Drosophila* APC tumour suppressor homolog. *BMC Cell Biol.* **5**, 37 (2004).
- Adams, C. L., Chen, Y. T., Smith, S. J. & Nelson, W. J. Mechanisms of epithelial cell-cell adhesion and cell compaction revealed by high-resolution tracking of E-cadherin-green fluorescent protein. *J. Cell Biol.* **142**, 1105–1119 (1998).
- Cox, R. T., Kirkpatrick, C. & Peifer, M. Armadillo is required for adherens junction assembly, cell polarity, and morphogenesis during *Drosophila* embryogenesis. *J. Cell Biol.* **134**, 133–148 (1996).
- Muller, H. A. & Wieschaus, E. Armadillo, bazooka, and stardust are critical for early stages in formation of the zonula adherens and maintenance of the polarized blastoderm epithelium in *Drosophila*. *J. Cell Biol.* **134**, 149–163 (1996).

13. Uemura, T. *et al.* Zygotic *Drosophila* E-cadherin expression is required for processes of dynamic epithelial cell rearrangement in the *Drosophila* embryo. *Genes Dev.* **10**, 659–671 (1996).
14. Harris, T. J. & Peifer, M. The positioning and segregation of apical cues during epithelial polarity establishment in *Drosophila*. *J. Cell Biol.* **170**, 813–823 (2005).
15. Tepass, U. & Hartenstein, V. The development of cellular junctions in the *Drosophila* embryo. *Dev. Biol.* **161**, 563–596 (1994).
16. Oda, H., Tsukita, S. & Takeichi, M. Dynamic behaviour of the cadherin-based cell-cell adhesion system during *Drosophila* gastrulation. *Dev. Biol.* **203**, 435–450 (1998).
17. Adams, C. L. & Nelson, W. J. Cyto mechanics of cadherin-mediated cell-cell adhesion. *Curr. Opin. Cell Biol.* **10**, 572–577 (1998).
18. Wiedenmann, J. *et al.* EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc. Natl Acad. Sci. USA* **101**, 15905–15910 (2004).
19. Kofron, M., Spagnuolo, A., Klymkowsky, M., Wylie, C. & Heasman, J. The roles of maternal α -catenin and plakoglobin in the early *Xenopus* embryo. *Development* **124**, 1553–1560 (1997).
20. Torres, M. *et al.* An α -E-catenin gene trap mutation defines its function in preimplantation development. *Proc. Natl Acad. Sci. USA* **94**, 901–906 (1997).
21. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
22. Pilot, F., Philippe, J.-M., Lemmers, C. & Lecuit, T. Spatial control of actin organization at adherens junctions by the synaptotagmin-like protein Btsz. *Nature* **442**, 580–584 (2006).
23. Kusumi, A., Sako, Y. & Yamamoto, M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophys. J.* **65**, 2021–2040 (1993).
24. Saxton, M. J. Single-particle tracking: the distribution of diffusion coefficients. *Biophys. J.* **72**, 1744–1753 (1997).
25. Somogyi, K. & Rorth, P. Evidence for tension-based regulation of *Drosophila* MAL and SRF during invasive cell migration. *Dev. Cell* **7**, 85–93 (2004).
26. Romero, S. *et al.* Formin is a processive motor that requires profilin to accelerate actin assembly and associated ATP hydrolysis. *Cell* **119**, 419–429 (2004).
27. Bogdan, S., Stephan, R., Lobke, C., Mertens, A. & Klambt, C. Abi activates WASP to promote sensory organ development. *Nature Cell Biol.* **7**, 977–984 (2005).
28. Kumar, S. *et al.* Viscoelastic retraction of single living stress fibers and its impact on cell shape, cytoskeletal organization, and extracellular matrix mechanics. *Biophys. J.* **90**, 3762–3773 (2006).
29. Dutta, D., Bloor, J. W., Ruiz-Gomez, M., VijayRaghavan, K. & Kiehart, D. P. Real-time imaging of morphogenetic movements in *Drosophila* using Gal4-UAS-driven expression of GFP fused to the actin-binding domain of moesin. *Genesis* **34**, 146–151 (2002).
30. Kametani, Y. & Takeichi, M. Basal-to-apical cadherin flow at cell junctions. *Nature Cell Biol.* **9**, 92–98 (2007).
31. Abe, K. & Takeichi, M. EPLIN mediates linkage of the cadherin catenin complex to F-actin and stabilizes the circumferential actin belt. *Proc. Natl Acad. Sci. USA* **105**, 13–19 (2008).
32. Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429**, 667–671 (2004).
33. Classen, A. K., Anderson, K. I., Marois, E. & Eaton, S. Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. *Dev. Cell* **9**, 805–817 (2005).
34. Izumi, G. *et al.* Endocytosis of E-cadherin regulated by Rac and Cdc42 small G proteins through IQGAP1 and actin filaments. *J. Cell Biol.* **166**, 237–248 (2004).
35. Kartenbeck, J., Schmelz, M., Franke, W. W. & Geiger, B. Endocytosis of junctional cadherins in bovine kidney epithelial (MDBK) cells cultured in low Ca²⁺ ion medium. *J. Cell Biol.* **113**, 881–892 (1991).
36. Troyanovsky, R. B., Sokolov, E. P. & Troyanovsky, S. M. Endocytosis of cadherin from intracellular junctions is the driving force for cadherin adhesive dimer disassembly. *Mol. Biol. Cell* **17**, 3484–3493 (2006).
37. Cavey, M. & Lecuit, T. *Imaging Cellular and Molecular Dynamics in Live Embryos Using Fluorescent Proteins* (ed. Dahmann, C.) (Humana, 2008).
38. Tsuji, A. & Ohnishi, S. Restriction of the lateral motion of band 3 in the erythrocyte membrane by the cytoskeletal network: dependence on spectrin association state. *Biochemistry* **25**, 6133–6139 (1986).
39. Qian, H., Sheetz, M. P. & Elson, E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* **60**, 910–921 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. M. Philippe for preparing the *E-cad::EosFP* construct, C. Picard for contribution to the α -Cat RNAi studies, and everyone who provided us with reagents, especially H. Oda, G. Nienhaus, D. Kiehart, C. Klambt, P. Rørth and E. Wieschaus. We also thank all members of the Lecuit and Lenne laboratories for discussions and comments on the manuscript. This work was supported by the CNRS, the Fondation Schlumberger pour l'Education et la Recherche (FSER), the ANR-Blanc together with P.-F.L.. M.C. was supported by the FSER and the Association pour la Recherche sur le Cancer (ARC). M.R. is supported by a Bourse Région PACA-Entreprise (Amplitude Systems).

Author Contributions T.L. planned the project and analysed the experiments together with M.C., M.R. and P.-F.L.; M.C. conducted the experiments except for the nano-ablations, which were performed by M.R.; P.-F.L. developed the nano-ablation system together with M.R. and the data quantification procedures with M.C. The manuscript was written by T.L. and all authors commented on it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.L. (lecuit@ibdm.univ-mrs.fr) or P.-F.L. (lenne@fresnel.fr).

METHODS

Fly strains. *w;ubi-E-cad::GFP* (homozygote on second chromosome; gift from H. Oda). *E-cad-GFP* is expressed at the same levels as endogenous *E-cad*, and rescues *E-cad* mutants⁴⁰. *w;UAS-DiaCA[ΔNAC]/TM6* (from P. Rørth), *w;CyO/Sp;UAS-Myr-Wsp/(TM2)* (from C. Klaemdt), *w;sqh-MoeABD-GFP* (from D. Kiehart), *w;Ubi-VSVG::GFP* (from T.L.). Overexpression experiments resulted from crosses of heterozygous females bearing the 67 driver (*mat αTub-gal4 VP16(67c)*) recombined with *ubi-E-cad::GFP* with males homozygous for *UAS-DiaCA* or *UAS-Myr-Wsp*.

The entire *E-cad* sequence (1,507 amino acids) was obtained from the Up-*E-cad::GFP* plasmid (from H. Oda), and mutagenized to replace the stop codon by a linker sequence coding for Ala-Met-Ser and corresponding to an NcoI site. EosFP (from G. Nienhaus) was subcloned carboxy-terminal to *E-cad* using NcoI and the resulting sequence was introduced into a pUASGateway vector (from T. Murphy). *E-cad-EosFP* was visualized in embryos from *67/UASp-E-cad::EosFP* mothers crossed to males homozygous for *UASp-E-cad::EosFP*.

Embryo injections. Embryo injection was performed as described²². Lat-A was injected at 1 mM (in 6% DMSO, 94% H₂O) yielding a ~20 μM concentration in embryos. Control DMSO injections had no effect on the stability of the epithelium. Rok inhibitor (Y27632) was injected at 1.13 mg ml⁻¹. For the RNAi injections, the *btsz* RNAi probe was made between nucleotides 3,065–3,556 of *btsz* sequence (*btsz2* transcript, GenBank accession AY229970) using the primers TAATACGACTCACTATAGGGAGACCACTAACAACAACGACGGCGAACTTG and TAATACGACTCACTATAGGGAGACCAACAATCGCCAGCTGCAC-TTTCATTG (underlined sequences represent the T7 promoter); the *α-Cat* RNAi probe was made between nucleotides 101–828 of *α-Cat* sequence (GenBank accession D13964) using the primers TAATACGACTCACTATA-GGGAGACCACCACAATGTCAGTTGAAAAACACTTG and TAATACGACTCACTATAGGGAGACCACTGTTGGGATGACTTTCCTTGGGCAAC. RNAi probes were injected at 5 μM (in H₂O) in freshly laid embryos.

Immunofluorescence. Embryos were fixed in 3.7% paraformaldehyde for 20 min and hand-devitellinized for *E-cad/F-actin* stainings, or post-fixed in methanol for *E-cad/β-Cat* stainings. Embryos were heat-fixed and post-fixed in methanol for *α-Cat/β-Cat* stainings. Injected embryos (Lat-A, *α-Cat* RNAi) were fixed and hand-devitellinized.

The antibodies we used were: rat anti-*E-cad* antibody (1:10; DCAD2 Hybridoma Bank), rat anti-*α-Cat* (1:10; DCAT1 Hybridoma Bank) and rabbit anti-*β-Cat* (1:250, gift from E. Wieschaus). Fixation and stainings were processed in the same batch and imaged with the same confocal settings. F-actin stainings were performed with Alexa 546 Phalloidin (Molecular Probes) at a 1:250 dilution for 20 min.

Quantification of *α-Cat* extinction in *α-Cat* RNAi-injected embryos (see Supplementary Fig. 3). For western blotting experiments, approximately 100–200 *α-Cat* RNAi or water-injected embryos were aged and collected from the coverslips in heptane, rinsed in PBS 0.1% Triton, snap frozen on dry ice, ground in embryo lysis buffer, centrifuged for 10 min at 13,000 r.p.m., and the supernatant was then boiled for 10 min with 5× loading buffer. Anti-*α-Cat* antibodies (DCAT1) were used at a 1:100 dilution and anti-*α-Tub* at 1:5,000 (Sigma). Signal detection was performed using an ECL UV imager (Fisher), and quantification of signal intensity using the Bio1D software (Vilber Lourmat).

Increasing concentrations of Oregon R embryo extracts were used to determine the slope of ECL signal versus protein concentration. This allowed us to determine how much *α-Cat* and *α-Tub* protein was present per lane for the *α-Cat* RNAi and water samples. *α-Cat* levels were then normalized to *α-Tub* levels.

Live imaging. Embryos were prepared and imaged as described³⁷. For *E-cad-GFP* FRAP and spot tracking, images were acquired using a Zeiss LSM510 confocal microscope, with a Plan-Apochromat 100× oil objective (numerical aperture of 1.4), and an Argon laser (488 nm). Each image was 512 × 300 pixels (1 pixel = 40 nm), averaged 4 times (line average) per time point, with a 921.6 ms scan time per image, and without delay between consecutive images. Photobleaching (FRAP) was performed with the Argon laser at full power, over an ROI with 21 pixel diameter (ROI radius = 420 nm; area of ROI = 0.554 μm²). Less than 80% of the initial fluorescence in the ROI was eliminated by photobleaching to ensure a gaussian distribution of molecules after bleaching⁴¹. Fluorescence recovery was then recorded during 45 or 63 s.

Photoconversion of *E-cad-EosFP* was performed with an ultraviolet laser (405 nm), acousto-optical tunable filter at 10%, typically 50–80 iterations (pixel time was 1.60 μsec⁻¹) over ROIs of variable size. The objective used was 40× water immersion, numerical aperture of 1.4. Non-converted *E-cad-EosFP* was excited at 488 nm and photoconverted *E-cad-EosFP* at 543 nm.

Nano-ablation. The ablation system was built from an epifluorescence inverted microscope (Axiovert 135M, Zeiss). A 50 MHz near-infrared (NIR) laser delivering femtosecond pulses at 1,030 nm (t-pulse, Amplitude Systems) was tightly focused by a high numerical aperture (1.2) 63× objective (C-Apochromat, Zeiss) transmitting 65% of the incoming NIR light.

Targeted junctions were exposed to NIR-femtosecond pulses between 3 and 6 ms with an average power of 370 mW at the back aperture of the objective. Similar set-ups were shown to allow ablation of subcellular structures below the diffraction limit⁴². The whole system was automated using implemented LABView (National Instruments) software. Classical epifluorescent illumination and a cooled CCD camera (DX2H, Kappa) were used. Images were processed using ImageJ (W. S. Rasband: <http://rsb.info.nih.gov/ij/>) software.

Nano-ablation was performed in the ventral-lateral region of embryos. To ensure that the plasma membrane had not been ruptured by laser ablation, we injected CMNB-caged fluorescein (fluorescein bis-(5-carboxymethoxy-2-nitrobenzyl) ether; Molecular Probes) into the embryo during cellularization. The fluorescent molecule was locally uncaged through a 3-photon process using the same NIR-femtosecond laser source that was used for ablation. The average power used to uncage was very low (40 mW) at the back aperture of the objective. Cell exposure to NIR-femtosecond laser light for uncaging ranged between 10 and 20 s.

FRAP analyses, *E-cad* cluster mobility and distribution relative to F-actin. Detailed information can be found in Supplementary Methods 7.

40. Oda, H. & Tsukita, S. Real-time imaging of cell-cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114**, 493–501 (2001).
41. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055–1069 (1976).
42. König, K. Multiphoton microscopy in life sciences. *J. Microsc.* **200**, 83–104 (2000).

High-harmonic generation by resonant plasmon field enhancement

Seungchul Kim^{1*}, Jonghan Jin^{1*}, Young-Jin Kim¹, In-Yong Park¹, Yunseok Kim¹ & Seung-Woo Kim¹

High-harmonic generation by focusing a femtosecond laser onto a gas is a well-known method of producing coherent extreme-ultraviolet (EUV) light^{1–3}. This nonlinear conversion process requires high pulse intensities, greater than $10^{13} \text{ W cm}^{-2}$, which are not directly attainable using only the output power of a femtosecond oscillator. Chirped-pulse amplification enables the pulse intensity to exceed this threshold by incorporating several regenerative and/or multi-pass amplifier cavities in tandem^{4,5}. Intracavity pulse amplification (designed not to reduce the pulse repetition rate) also requires a long cavity^{6,7}. Here we demonstrate a method of high-harmonic generation that requires no extra cavities. This is achieved by exploiting the local field enhancement induced by resonant plasmons within a metallic nanostructure consisting of bow-tie-shaped gold elements on a sapphire substrate. In our experiment, the output beam emitted from a modest femtosecond oscillator (100-kW peak power, 1.3-nJ pulse energy and 10-fs pulse duration) is directly focused onto the nanostructure with a pulse intensity of only $10^{11} \text{ W cm}^{-2}$. The enhancement factor exceeds 20 dB, which is sufficient to produce EUV wavelengths down to 47 nm by injection with an argon gas jet. The method could form the basis for constructing laptop-sized EUV light sources for advanced lithography and high-resolution imaging applications.

Field enhancement is attributed to the collective motion of free electrons confined in narrowly localized regions, similar to that

observed in colloidal nanoparticles exposed to an external electromagnetic field^{8,9}. For the given geometry of a nanoparticle or nanostructure, the degree of field enhancement induced by resonant plasmons can be estimated by solving Maxwell's equations analytically¹⁰ or numerically¹¹. Using appropriate fabrication techniques, it is consequently possible to optimize the shape of a nanostructure to be tailored to a particular application^{12–15}. We adopted this approach in the present investigation, to design and build a nanostructure that enables the generation of high harmonics from a modest femtosecond oscillator through field enhancement around the nanostructure.

Figure 1 illustrates the construction of the experimental apparatus. The femtosecond oscillator used here is a titanium-sapphire oscillator (Femtosecond sPRO, Femtolasers) set to emit a train of 10-fs pulses of 800-nm carrier wavelength at a repetition rate of 75 MHz. The output beam from the oscillator has a small peak power of $\sim 100 \text{ kW}$ and each pulse has an energy of 1.3 nJ, yielding a pulse intensity of only $\sim 10^{11} \text{ W cm}^{-2}$ even when well focused. The pulse intensity must be increased at least by two orders of magnitude to reach the threshold required to generate high harmonics by interaction with a gas jet. To do this, a metallic nanostructure consisting of a two-dimensional array of gold 'bow tie' elements on a sapphire plate is inserted in the focal plane of the focused beam.

The pulse intensity builds up when the femtosecond pulse passes through the nanostructure. The degree of field enhancement is

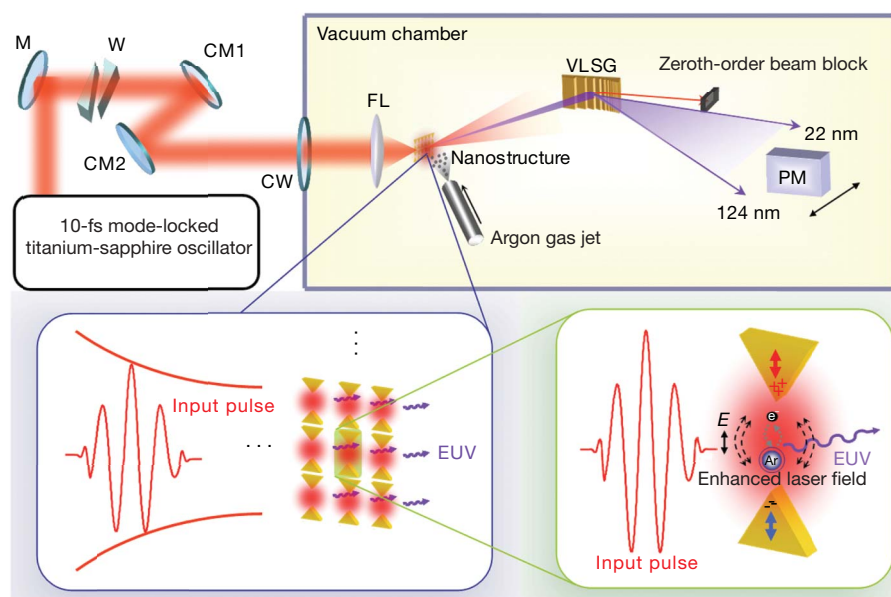


Figure 1 | Apparatus for high-harmonic generation by electric field enhancement using a nanostructure of bow-tie elements.

Anticlockwise from top: the overall system configured for the experiment, a detailed view of the two-dimensional array of bow-tie elements fabricated on a sapphire plate, and the magnified view of a single gold bow-tie element interacting with the incident pulse. Ar, argon atom; CM, chirped mirror; CW, chamber window; FL, focusing lens; M, mirror; PM, photon multiplier; VLSG, varied-line-spacing grating; W, wedge plate; E, electric field.

¹Billionth Uncertainty Precision Engineering Group, KAIST, Daedeok Science Town, Daejeon 305-701, South Korea.

*These authors contributed equally to this work.

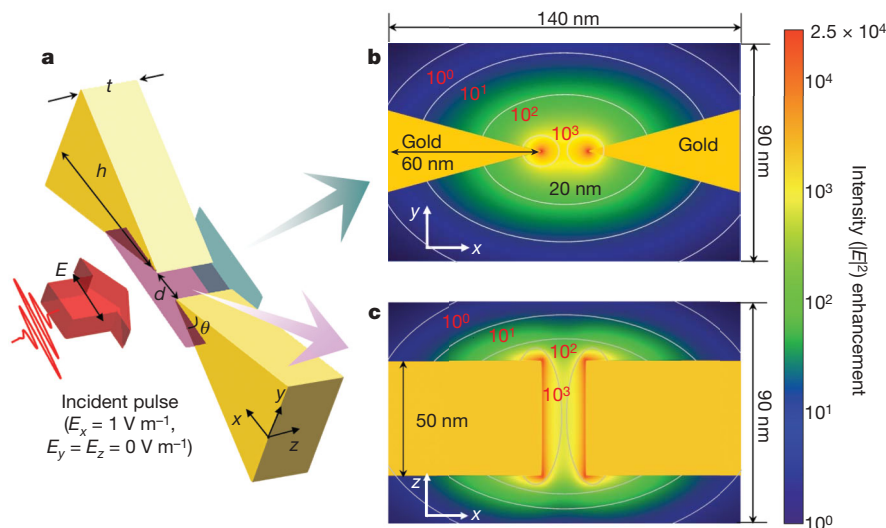


Figure 2 | Finite-difference time-domain simulation of local field enhancement. **a**, Geometrical conditions for simulation of a single bow-tie element. The polarization direction of the incident pulse parallel to the x - z plane. **b**, Intensity field computed in the x - y plane. **c**, Same simulation result

significantly affected by the geometrical shape of the nanostructure, as revealed in previous investigations of different shapes of particles, wires, and tips^{16–19}. The bow-tie shape of the nanostructure elements—a pair of triangular patches placed apex to apex with a small gap between them—is taken as the nanostructure's basic feature. When exposed to the femtosecond pulse, free electrons are confined within one apex of the bow-tie element and the opposing apex is filled with positive charges, which makes the pulse capable of producing a strong field enhancement between the vertices.

As illustrated in Fig. 2, the bow-tie shape is characterized by four geometrical parameters, the thickness (t), angle (θ), gap (d) and height (h). These parameters must be selected so as to maximize the resulting field enhancement. For this, the finite-difference time-domain method was adopted^{20,21}, and Maxwell's equations for the case shown in Fig. 2a were numerically solved. The polarization direction of the incident femtosecond pulse is positioned parallel to the line connecting the vertices. The incident pulse is assumed to have a bandwidth of 100 nm about the 800-nm carrier wavelength. The wavelength-dependent dielectric constant of the tip material, gold, was determined using the modified Debye model²². Through trial-and-error simulation, the four parameters were finally determined as follows: $h = 175$ nm, $d = 20$ nm, $t = 50$ nm and $\theta = 30^\circ$. The corresponding field enhancement was computed as shown in Figs 2b and 2c. The entire volume of simulation, including a bow-tie element, was evenly divided into hexahedral grids $1\text{ nm} \times 1\text{ nm} \times 1\text{ nm}$ in size. No singularity effect along the edge line of the element was considered. The computation shows that the intensity enhancement factor reaches its maximum, 44 dB, at the vertices. In addition, within the $60\text{ nm} \times 50\text{ nm} \times 50\text{ nm}$ ($x \times y \times z$) region in the gap between the vertices, the enhancement factor is greater than 20 dB. This result confirms that the field enhancement suffices to boost the pulse intensity to well above the required threshold of 10^{13} W cm^{-2} , given that the input pulse has an intensity of 10^{11} W cm^{-2} .

The time-dependent dynamic behaviour of plasmon resonance within the nanostructure causes a temporal phase delay in the enhanced field in response to the original pulse. The simulation (Fig. 2) revealed a uniform phase delay of 90° , confirming the phase consistency over the entire enhanced field. In addition to gold, three other metals, silver, copper and platinum, were also considered. However, simulation showed that no significant difference existed in the resulting field enhancement, other than in platinum being less effective. Thus, we chose gold as the bow-tie

viewed in the x - z plane. The design parameters selected were $h = 175$ nm, $d = 20$ nm, $t = 50$ nm and $\theta = 30^\circ$. The intensity enhancement factor, which reached a maximum of 2.5×10^4 at the apex of each triangle, turned out to be $>10^2$ within the $60\text{ nm} \times 50\text{ nm} \times 50\text{ nm}$ ($x \times y \times z$) volume in the gap.

material in consideration of our experience with the metal in nanofabrication.

Figure 3 shows a scanning electron microscope image of the nanostructure actually fabricated on a sapphire plate of $400\text{-}\mu\text{m}$ thickness. Using the electron beam evaporation process, a 50-nm gold layer was deposited on a 5-nm chromium adhesion layer on the sapphire plate. By means of precise control of the focused ion beam, we then made bow-tie elements by scribing the deposited gold and chromium layers with subnanometer lateral resolution. The sapphire plate serves as a solid transparent substrate holding bow-tie elements in a two-dimensional array with a spatial pitch of 200 nm in one direction and 550 nm in the perpendicular direction, over a $10\text{ }\mu\text{m} \times 10\text{ }\mu\text{m}$ area. The sapphire plate acts as a heat dissipator that protects the nanostructure from thermal damage both during nanofabrication

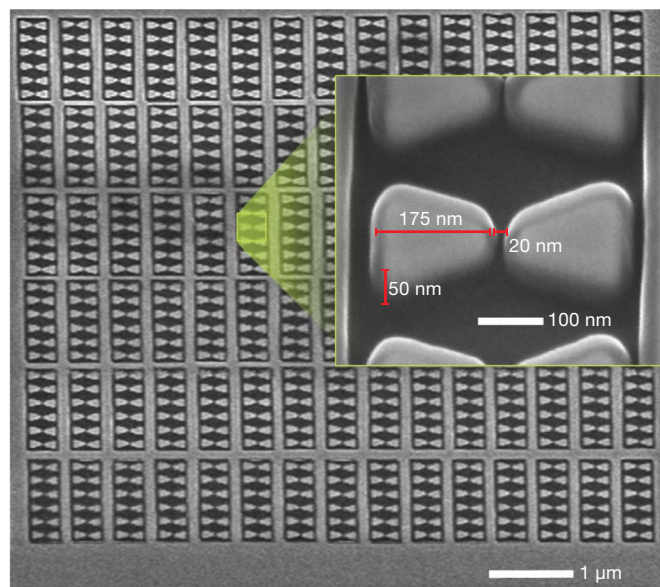


Figure 3 | Scanning electron microscope image of the nanostructure used for high-harmonic generation. Bow-tie elements were arranged in a two-dimensional, 36×15 array with an area of $10\text{ }\mu\text{m} \times 10\text{ }\mu\text{m}$. The inset shows the magnified image of a single bow-tie element with the important dimensions marked. Owing to the high magnification, edge lines are seen blurred by multiple scattering of electrons in imaging.

and when it is illuminated by the femtosecond laser for high-harmonic generation.

As shown in Fig. 1, the sapphire plate was placed inside the vacuum chamber and the nanostructure was positioned to face the nozzle of the gas jet. The output beam of the femtosecond laser was fed into the vacuum chamber through a window and then focused on the nanostructure using an aspheric focusing lens. The group delay dispersion, which was caused by the chamber window, the focusing lens and the sapphire plate, and amounted to 400 fs^2 , was compensated for using chirped mirrors together with a pair of wedges. Inside the chamber, a vacuum of 10^{-4} – 10^{-5} torr was maintained by operating a rotary pump together with a diffusion pump equipped with a liquid nitrogen trap. The apparatus was placed on a granite table floated on air supports to prevent external vibration. A gas jet of argon was injected through a nozzle of $100\text{-}\mu\text{m}$ diameter to provide a surface density of 0.29 g per litre by a back pressure of 115 torr .

To detect the generated high harmonics, a high-resolution spectrometer was constructed by combining a type of diffraction grating with varied line spacing (Hitachi, 001-0639) with a photon multiplier (Photonics, 4751G CSI). The photon multiplier was located at a distance of 469 mm from the grating, which dispersed the incoming EUV radiation in the $22\text{--}124\text{-nm}$ wavelength range along a transverse line 110 mm long. The photon multiplier was mounted on a precision stage moving along the dispersion line of the grating. The photon multiplier we used was sensitive to high-harmonic radiation only below 200 nm in wavelength; hence, no aluminium filters were required to block the original femtosecond laser beam. The grating has a narrow acceptance angle of $\pm 1.0^\circ$, which provides a 1.0-nm wavelength sampling resolution. In addition, a slit of 1.0-mm width was installed to limit the aperture size of the photon multiplier to avoid the effect of excessive convolution in the measured data.

Figure 4 shows the detection result obtained in this investigation. Odd harmonics up to the 17th, denoted H17, can be clearly seen. H3 and H5 are not seen, because they are below the lower limit of the dispersion range of the grating we used. The three-stage power variation usually seen in the harmonics generated by chirped power amplification was also observed^{23,24}, as evidenced by a perturbative region from H7 to H11, a plateau from H11 to H15, and a final cutoff

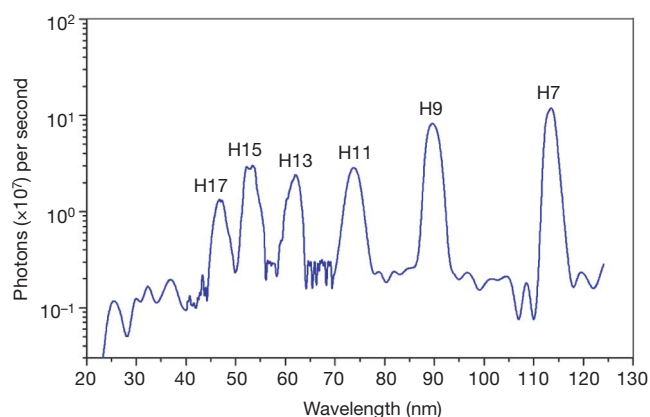


Figure 4 | Measured spectrum of generated high harmonics. A varied-line-spacing grating was used to disperse the $22\text{--}124\text{-nm}$ wavelength band over a total linear length of 110 mm . H7, H9 and H11 were monitored by moving a photon multiplier in steps of 1.0 mm along the dispersion line, and H13, H15 and H17 were monitored in steps of 0.2 mm . The aperture of the photon multiplier was adjusted to be 1.0 mm to reduce the unwanted convolution effect in the measured spectrum while maintaining a desirable level of photon-counting sensitivity. Given the input pulse power of 100 mW , the efficiencies for the harmonics were computed as follows: H7, 2.4×10^{-9} ; H9, 2.2×10^{-9} ; H11, 9.5×10^{-10} ; H13, 9.0×10^{-10} ; H15, 1.3×10^{-9} ; and H17, 6.9×10^{-10} .

at H17. Given that the input femtosecond laser provides a small peak power of 100 kW , the observed harmonics yield relatively low power, in the range of a few nanowatts in total. However, the conversion efficiency was found to be on the level of 2.4×10^{-9} for H7 and as low as 6.9×10^{-10} for H17, which is comparable to those of chirped power amplification. We expect that the original comb structure is maintained in each of the generated harmonics as there was no reduction in the repetition rate of the incident femtosecond pulses, although experimental verification of this is not within the scope of the current investigation.

Sapphire could break down owing to multiphoton ionization if it were exposed to an intensity of greater than $10^{13} \text{ W cm}^{-2}$ (ref. 25). However, the sapphire substrate used in the experiment was found not to be substantially influenced, because the electric field exceeding the threshold intensity was concentrated within the gap between the triangular electrodes. This was verified by scanning electron microscope imaging after the experiment. The gold electrodes themselves are also susceptible to thermal damage owing to photon–electron coupling when the incident femtosecond pulse yields an intensity of greater than $10^{11} \text{ W cm}^{-2}$ (ref. 26). Because the sapphire substrate functions as a good heat dissipator, no serious ablation or melting of the electrodes was observed, however, as long as the incident intensity was not allowed to exceed the threshold intensity.

Another concern was the spatial distribution shared by the high harmonics generated using the array of bow-tie elements. The intense spot of the incident femtosecond pulse focused on the surface of the nanostructure had a diameter of $5 \mu\text{m}$, which is large enough to produce high-harmonic generation simultaneously in approximately 150 of the elements. Each element acts as a point-like source that radiates high harmonics with a broad angle distribution. The consequent wave collectively observed from the elements is therefore the constructive interference of the individual waves emitted from the individual elements, which maintain high coherence in both spatial and temporal terms. This causes the generated harmonics to be distributed following the particular spatial diffraction pattern that the nanostructure would produce as a two-dimensional diffractive grating as a whole.

For instance, H17, which has a wavelength of 47 nm , has a first-order diffraction angle of 13.6° in one direction and 4.9° in the perpendicular direction, with greater higher-order diffraction. In the experiment, only the zeroth-order diffracted ray of H17 was collected by the photon multiplier, because the other diffraction angles are greater than the acceptance angle of $\pm 1.0^\circ$ of the spectrometer used. This is the case for the other detected harmonics, H7–H15, as well. The zeroth-order diffracted rays of the generated harmonics all have the same propagation direction as the incident femtosecond pulse, and are thus confined within a narrow cone with a $\pm 7^\circ$ angle distribution formed in response to the focusing angle of the incident pulse.

A final concern was the dispersion-induced phase mismatching due to the interaction of the generated harmonics with the injected argon gas. However, we did not pay particular attention to this problem, because the field enhancement for the high-harmonic generation was confined to region of dimension less than $1 \mu\text{m}$ as the incident femtosecond pulse passed through the bow-tie elements.

In conclusion, in the experiment performed in this study we successfully verified that the field enhancement induced around the bow-tie elements with a 20-nm gap allows the generation of EUV light directly from the output of a single femtosecond oscillator of 100-kW peak power. The focused pulse intensity of $10^{11} \text{ W cm}^{-2}$ was enhanced by a factor of $20\text{--}40 \text{ dB}$, which is large enough to produce harmonics as high as the 17th, corresponding to a wavelength of 47 nm . Requiring no extra cavities, the proposed method enables the construction of convenient laptop-size sources of coherent EUV radiation. This has the potential to accelerate progress in many areas of science and technology, particularly in the area of advanced

lithography, high-resolution imaging and possibly EUV optical clocks.

Received 28 January; accepted 11 April 2008.

1. Corkum, P. B. Plasma perspective on strong-field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
2. Lewenstein, M., Balcou, P., Ivanov, M., Yu, L'Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
3. Chang, Z., Rundquist, A., Wang, H., Murnane, M. M. & Kapteyn, H. C. Generation of coherent soft X rays at 2.7 nm using high harmonics. *Phys. Rev. Lett.* **79**, 2967–2970 (1997).
4. Strickland, D. & Mourou, G. Compression of amplified chirped optical pulses. *Opt. Commun.* **56**, 219–221 (1985).
5. Seres, J. *et al.* Laser technology: Source of coherent kiloelectronvolt X-rays. *Nature* **433**, 596 (2005).
6. Gohle, C. *et al.* A frequency comb in the extreme ultraviolet. *Nature* **436**, 234–237 (2005).
7. Jones, R. J., Moll, K. D., Thorpe, M. J. & Ye, J. Phase-coherent frequency combs in the vacuum ultraviolet via high-harmonic generation inside a femtosecond enhancement cavity. *Phys. Rev. Lett.* **94**, 193201 (2005).
8. Nisoli, M. Nanoplasmonics: Brave new attoworld. *Nature Photon.* **1**, 499–500 (2007).
9. Roth, M. R., Panoiu, C. N., Adams, M. M. & Osgood, M. R. Jr. Resonant-plasmon field enhancement from asymmetrically illuminated conical metallic-probe tips. *Opt. Express* **14**, 2921–2931 (2006).
10. Schatz, G. C. & Van Duyne, R. P. *Handbook of Vibrational Spectroscopy* Vol. 1 (eds Chalmers, J. M. & Griffiths, P. R.) 759–774 (Wiley, Chichester, 2002).
11. Bohn, L. J., Nesbitt, J. D. & Gallagher, A. Field enhancement in apertureless near-field scanning optical microscopy. *J. Opt. Soc. Am. A* **18**, 2998–3006 (2001).
12. Mühlischlegel, P., Eisler, H.-J., Martin, O. J. F., Hecht, B. & Pohl, D. W. Resonant optical antennas. *Science* **308**, 1607–1609 (2005).
13. Bragas, A. V. & Martínez, O. E. Field-enhanced scanning optical microscope. *Opt. Lett.* **25**, 631–633 (2000).
14. Stockman, M. I. Nanofocusing of optical energy in tapered plasmonic waveguides. *Phys. Rev. Lett.* **93**, 137404 (2004).
15. Lesuffleur, A., Kumar, L. K. S. & Gordon, R. Enhanced second harmonic generation from nanoscale double-hole arrays in a gold film. *Appl. Phys. Lett.* **88**, 261104 (2006).
16. Jin, E. X. & Xu, X. Enhanced optical near field from a bowtie aperture. *Appl. Phys. Lett.* **88**, 153110 (2006).
17. Onuta, T.-D., Waagele, M., DuFort, C. C., Schaich, W. L. & Dragnea, B. Optical field enhancement at cusps between adjacent nanoapertures. *Nano Lett.* **7**, 557–564 (2007).
18. Sun, W.-X. & Shen, Z.-X. Optimizing the near field around silver tips. *J. Opt. Soc. Am. A* **20**, 2254–2259 (2003).
19. Tzeng, Y.-F. *et al.* Fabrication of an ultra-nanocrystalline diamond-coated silicon wire array with enhanced field-emission performance. *Nanotechnology* **18**, 435703 (2007).
20. Fromm, D. P., Sundaramurthy, A., James Schuck, P., Kino, G. & Moerner, W. E. Gap-dependent optical coupling of single “bowtie” nanoantennas resonant in the visible. *Nano Lett.* **4**, 957–961 (2004).
21. Krug, J. T. II, Sánchez, E. J. & Xie, X. S. Design of near-field optical probes with optimal field enhancement by finite difference time domain electromagnetic simulation. *J. Chem. Phys.* **116**, 10895–10901 (2002).
22. Gai, H., Wang, J. & Tian, Q. Modified Debye model parameters of metals applicable for broadband calculations. *Appl. Opt.* **46**, 2229–2233 (2007).
23. Zhou, J., Peatross, J., Murnane, M. M., Kapteyn, H. C. & Christov, I. P. Enhanced high-harmonic generation using 25 fs laser pulses. *Phys. Rev. Lett.* **76**, 752–755 (1996).
24. Wahlström, C.-G. *et al.* High-order harmonic generation in rare gases with an intense short pulse laser. *Phys. Rev. A* **48**, 4709–4720 (1993).
25. Wang, X. C. *et al.* Femtosecond pulse laser ablation of sapphire in ambient air. *Appl. Surf. Sci.* **228**, 221–226 (2004).
26. Güdde, J., Hohlfeld, J., Müller, J. G. & Matthias, E. Damage threshold dependence on electron-phonon coupling in Au and Ni films. *Appl. Surf. Sci.* **127–129**, 40–45 (1998).

Acknowledgements This work was supported by the Creative Research Initiatives Program of the Ministry of Science and Technology in the Republic of Korea.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.-W.K. (swk@kaist.ac.kr).

Superconductivity at 43 K in $\text{SmFeAsO}_{1-x}\text{F}_x$

X. H. Chen¹, T. Wu¹, G. Wu¹, R. H. Liu¹, H. Chen¹ & D. F. Fang¹

Since the discovery of high-transition-temperature (high- T_c) superconductivity in layered copper oxides, extensive effort has been devoted to exploring the origins of this phenomenon. A T_c higher than 40 K (about the theoretical maximum predicted from Bardeen–Cooper–Schrieffer theory¹), however, has been obtained only in the copper oxide superconductors. The highest reported value for non-copper-oxide bulk superconductivity is $T_c = 39$ K in MgB_2 (ref. 2). The layered rare-earth metal oxypnictides LnOFeAs (where Ln is La–Nd, Sm and Gd) are now attracting attention following the discovery of superconductivity at 26 K in the iron-based $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ (ref. 3). Here we report the discovery of bulk superconductivity in the related compound $\text{SmFeAsO}_{1-x}\text{F}_x$, which has a ZrCuSiAs-type structure. Resistivity and magnetization measurements reveal a transition temperature as high as 43 K. This provides a new material base for studying the origin of high-temperature superconductivity.

$\text{LnO}_{1-x}\text{F}_x\text{FeAs}$ adopts ZrCuSiAs-type structure. A series of equiatomic quaternary compounds LnFeAsO and LnFePO (Ln = La–Nd, Sm, Gd) with ZrCuSiAs-type structure has been reported^{4,5}. The crystal structure of the tetragonal ZrCuSiAs-type compound $\text{SmFeAs}(\text{O},\text{F})$ is shown in Fig. 1.

Polycrystalline samples with nominal composition $\text{SmFeAsO}_{1-x}\text{F}_x$ ($x = 0.15$) were synthesized by conventional solid state reaction using high-purity SmAs, SmF_3 , Fe and Fe_2O_3 as starting materials. SmAs was obtained by reacting Sm chips and As pieces at 600 °C for 3 h and then 900 °C for 5 h. The raw materials were thoroughly ground and pressed into pellets. The pellets were wrapped in Ta foil, sealed in an evacuated quartz tube, and finally annealed at either 1,160 °C or 1,200 °C for 40 h.

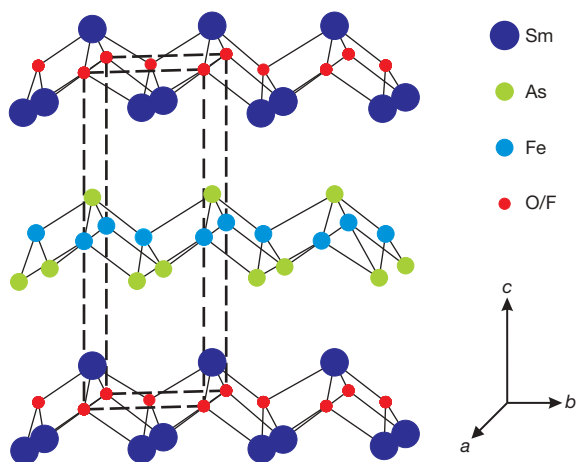


Figure 1 | Structural model of $\text{SmFeAsO}_{1-x}\text{F}_x$ with the tetragonal ZrCuSiAs-type structure. The quaternary equiatomic ZrCuSiAs-type structure is very simple, with only eight atoms in the tetragonal cell. The dashed lines represent a unit cell.

Figure 2 shows the X-ray diffraction (XRD) pattern for a sample annealed at 1,160 °C. It is found that the peaks in the XRD pattern can be well indexed to the tetragonal ZrCuSiAs-type structure with $a = 0.3932$ nm and $c = 0.8490$ nm, except for some tiny peaks from the impurity phase SmOF. These lattice parameters are slightly smaller than the values of $a = 0.3940$ nm and $c = 0.8501$ nm for F-free SmFeAsO .

Magnetic characterization of the superconducting transitions under a magnetic field of 10 Oe for a sample annealed at 1,160 °C is shown in Fig. 3; data are shown for the zero-field cooled and field-cooled measurements. The magnetic onset for the superconducting transition is 41.8 K for the sample annealed at 1,160 °C, and 41.3 K for the sample annealed at 1,200 °C (not shown). The existence of the superconducting phase was unambiguously confirmed by the Meissner effect on cooling in a magnetic field. A superconducting volume fraction of about 50% under a magnetic field of 10 Oe was obtained at 5 K, indicating that the superconductivity is bulk in nature.

Figure 4 shows the temperature dependence of the resistivity under zero magnetic field and under fields of $H = 5$ and 7 T; Fig. 4a and b show data for samples annealed at 1,160 °C and 1,200 °C, respectively. Under zero magnetic field, the onset transition and midpoint temperatures of the resistive transition are respectively 43 K and 41.7 K for a sample annealed at 1,160 °C; for a sample annealed at

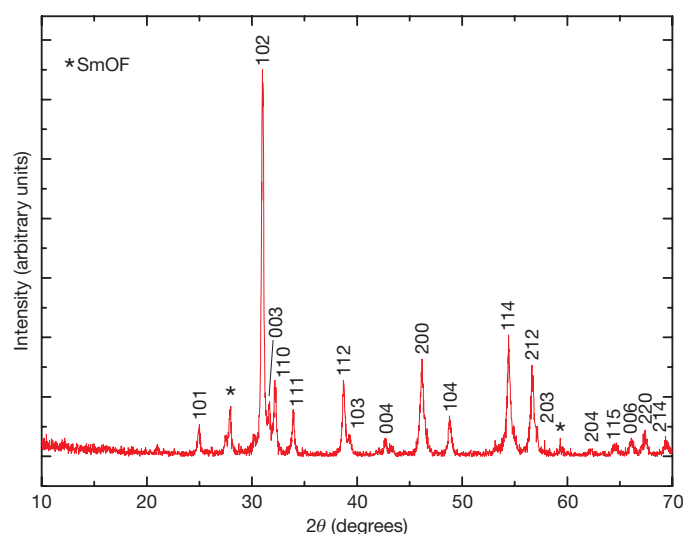


Figure 2 | X-ray diffraction pattern for a sample with nominal composition $\text{SmFeAsO}_{1-x}\text{F}_x$ ($x = 0.15$). The sample was annealed at 1,160 °C, and a tiny impurity phase SmOF is observed (stars denote peaks due to this impurity phase). The sample preparation process, except for annealing, was carried out in a glove box filled with a high-purity argon atmosphere. The samples were characterized at room temperature by X-ray diffraction using a Rigaku D/max-A X-ray diffractometer with $\text{Cu K}\alpha$ radiation ($\lambda = 0.15418$ nm) in the 2θ range of 10–70° with steps of 0.02°.

¹Hefei National Laboratory for Physical Sciences at Microscale and Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China.

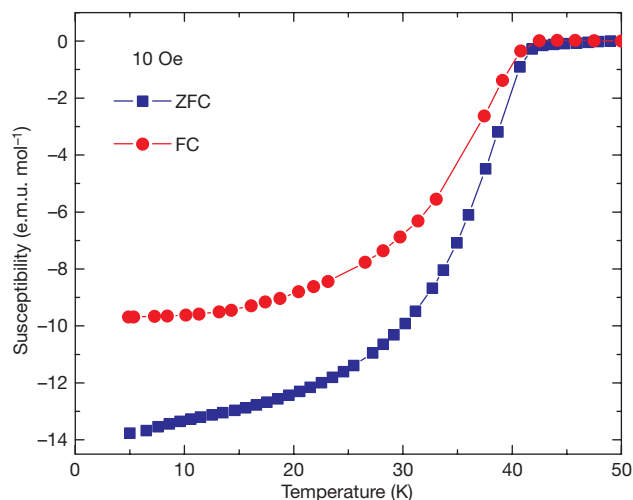


Figure 3 | Temperature dependence of magnetic susceptibility for a sample annealed at 1,160 °C. Data are shown for zero-field cooled (ZFC) and field-cooled (FC) measurements at 10 Oe. The susceptibility measurement was performed in an MPMS-7T system (Quantum Design).

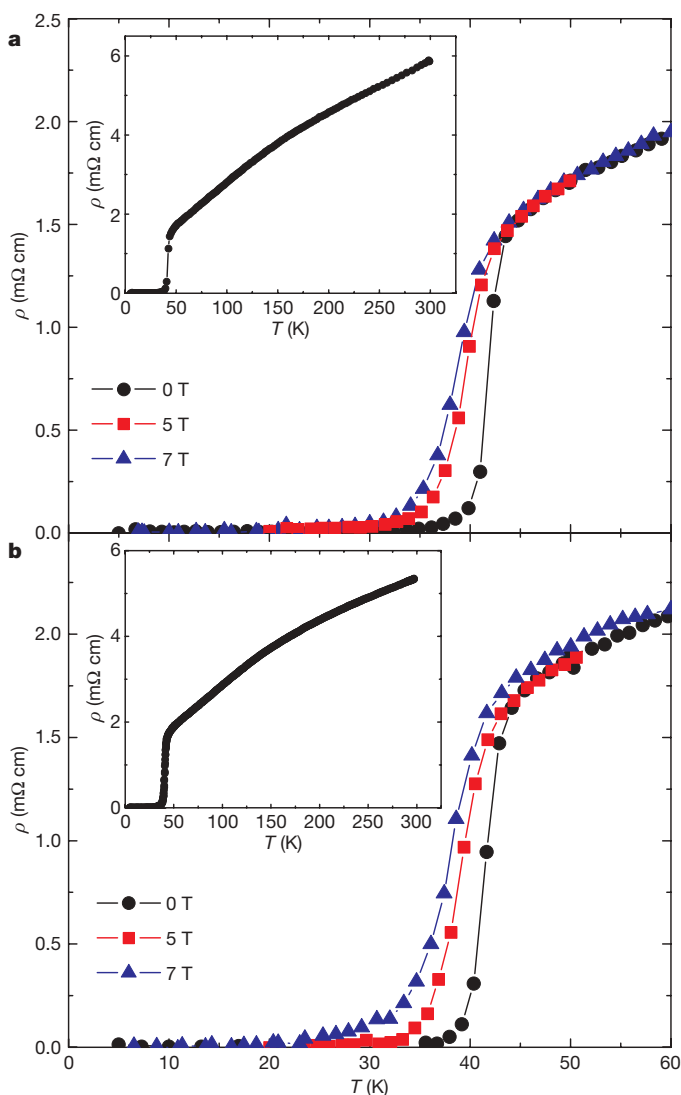


Figure 4 | Temperature dependence of resistivity with and without a magnetic field. **a**, Sample annealed at 1,160 °C; **b**, sample annealed at 1,200 °C. Insets, resistivity from 300 K to 5 K. Resistivity measurements were performed using an a.c. resistance bridge (Linear Research Inc., Model LR700) by the standard four-probe method. The transport properties were measured under magnetic fields of 5 and 7 T with an MPMS-7T system (Quantum Design).

1,200 °C, the values are respectively 43.7 K and 41.2 K. The 90–10% transition width are 2.5 K and 3 K for the samples annealed at 1,160 °C and 1,200 °C, respectively. It is found that the onset transition temperature in susceptibility coincides with the transition midpoint temperature in resistivity. An external magnetic field of 5 or 7 T makes the transition width broader, but the onset transition temperature is not sensitive to magnetic field, indicating that the upper critical field is very high for this superconductor. Therefore, this superconductor has potential applications due to its high transition temperature and high upper critical field.

Replacement of La by Sm leads to a large increase in T_c from 26 K in $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ (ref. 3) to 43 K in $\text{SmFeAsO}_{1-x}\text{F}_x$ (this work). This suggests that it is possible to realize higher T_c values in such layered oxypnictides. The observed T_c of 43 K in $\text{SmFeAsO}_{1-x}\text{F}_x$ is higher than the theoretical value predicted from Bardeen–Cooper–Schrieffer (BCS) theory¹, and this provides a strong argument for considering layered oxypnictide superconductors as unconventional superconductors.

Received 25 March; accepted 29 April 2008.

Published online 25 May 2008.

- McMillan, W. L. Transition temperature of strong-coupled superconductors. *Phys. Rev.* **167**, 331–344 (1968).
- Nagamatsu, J. *et al.* Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
- Kamihara, Y. *et al.* Iron-based layered superconductor $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ ($x=0.05\text{--}0.12$) with $T_c=26$ K. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
- Quebe, P. *et al.* Quaternary rare earth transition metal arsenide oxides RTAsO ($T=\text{Fe, Ru, Co}$) with ZrCuSiAs type structure. *J. Alloy. Comp.* **302**, 70–74 (2000).
- Zimmer, B. I. *et al.* The rare earth transition metal phosphide oxides LnFePO , LnRuPO and LnCoPO with ZrCuSiAs type structure. *J. Alloy. Comp.* **229**, 238–242 (1995).

Acknowledgements This work was supported by the Natural Science Foundation of China and by the Ministry of Science and Technology of China.

Author Contributions X.H.C. designed and coordinated the whole experiment, including the details of doping and synthesis, did some of the experiments, analysed the data and wrote the paper. T.W., G.W. and R.H.L. contributed equally to the synthesis, magnetic measurements and resistive measurements under magnetic field, H.C. did the structure analysis, and D.F.F. did some of the resistive measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.H.C. (chenxh@ustc.edu.cn).

Charge self-regulation upon changing the oxidation state of transition metals in insulators

Hannes Raebiger¹, Stephan Lany¹ & Alex Zunger¹

Transition-metal atoms embedded in an ionic or semiconducting crystal can exist in various oxidation states that have distinct signatures in X-ray photoemission spectroscopy and ‘ionic radii’ which vary with the oxidation state of the atom. These oxidation states are often tacitly associated with a physical ionization of the transition-metal atoms^{1,2}—that is, a literal transfer of charge to or from the atoms. Physical models have been founded on this charge-transfer paradigm^{3–6}, but first-principles quantum mechanical calculations show only negligible changes in the local transition-metal charge^{7–12} as the oxidation state is altered. Here we explain this peculiar tendency of transition-metal atoms to maintain a constant local charge under external perturbations in terms of an inherent, homeostasis-like negative feedback. We show that signatures of oxidation states and multivalence—such as X-ray photoemission core-level shifts, ionic radii and variations in local magnetization—that have often been interpreted as literal charge transfer^{3,4,13–16} are instead a consequence of the negative-feedback charge regulation.

Classic inorganic chemistry of transition-metal (TM) coordination compounds and semiconductors^{1,2} tacitly assumes that when the total charge of the compound is altered (for example through oxidation/reduction, carrier injection, chemical doping or decomposition of the compound), most of the changes are accommodated by a change in the charge of the TM ion. For example, when LiTMO₂ is delithiated, the original TM³⁺ ion is thought to be converted to a TM⁴⁺ ion. Although this concept of formal charges and changes therein provides useful bookkeeping, in many cases these formal charge states have been considered physical entities, capable of physical (for example point-ion Coulomb) interactions. Further examples of this concept include (1) the attribution³ of a spatial ordering of inequivalent Mn atoms in RMnO₃ manganites (where R represents La, ..., Tb or Dy) to Coulomb ordering of point charges associated with the formal charges of Mn⁴⁺ and Mn³⁺ ions; (2) the assumption that photoemission core-level shifts of TM ions within compounds reflect the physical charge of the TM ion itself^{3,4}; and (3) the view that the variation of TM–ligand bond length with oxidation state⁵ (‘ionic radii’) of TMs is a reflection of an explicit charge transfer between the TM and the ligand.

To address this concept of charge transfer on a general level, we specifically study multiple-charge configurations of isolated TM atoms in representative host materials. These include the archetypal ionic and covalent compounds, respectively MgO and GaAs, as well as an intermediate case involving *d* electrons, namely Cu₂O. We illustrate the negative charge feedback in terms of a simple energy-level model^{7,8,17} (Fig. 1). Here the isolated TM(*d*) orbitals occupied by *n* electrons split into crystal field levels with irreducible representations γ (for example, for cubic symmetry these levels have the representations $\gamma = t$ (triply degenerate) and $\gamma = e$ (doubly degenerate)). The crystal field levels TM(γ) (Fig. 1a, left) that have the same

representation and spin as the anion dangling bond levels DB(γ) of the cation vacancy (Fig. 1a, right) interact with these DB(γ) levels, and the interaction causes bonding levels to form below the valence band maximum and antibonding levels to form inside the band gap (Fig. 1a, centre). When the energy of the TM(γ) level is lower than that of the DB(γ) level, the bonding level occurs as a crystal field resonance (CFR) with a strong TM character and the antibonding level occurs as a dangling bond hybrid (DBH) formed mostly of the ligand *p* orbitals, and vice versa when the energy of the TM(γ) level is higher than that of the DB(γ) level, as illustrated in Fig. 1a, b.

Figure 1b shows how energy levels evolve when the system in Fig. 1a is doped with an electron. The level occupancy of the TM-induced hybrid states increases, and the system charge *q* becomes *q*′ = *q* − 1 (Fig. 1b). Such charge doping causes the TM levels and, consequently, the bonding and antibonding hybrid levels to shift up in energy by means of a negative feedback: in Fig. 1a the isolated TM(γ^n) orbital level (*n* denotes the electron occupancy) is lower in energy than the host DB(γ) level, so the bonding level occurs as a crystal field resonance, γ^{CFR} , and the antibonding level occurs as a dangling bond hybrid, γ^{DBH} , just above the valence band maximum. In Fig. 1b the atomic TM level, TM(γ^{n+1}), is higher in energy than the DB(γ) level, and the CFR occurs as the gap level whereas the DBH occurs as resonant state inside the host bands. Thus, the relative weight of the bonding levels shifts towards the ligands, and this negative feedback causes a depopulation of TM charge in the bonding states, counterbalancing the increase in the antibonding gap-level charge. As a result of this self-regulated response, the net local charge at the TM site remains approximately constant. In this model, the formal oxidation state does not reflect the local charge at the TM site, but rather the occupancy of the respective crystal field levels (for example $e^2 t^2$ for Mn_{Ga}⁰ in GaAs, see Supplementary Information section A).

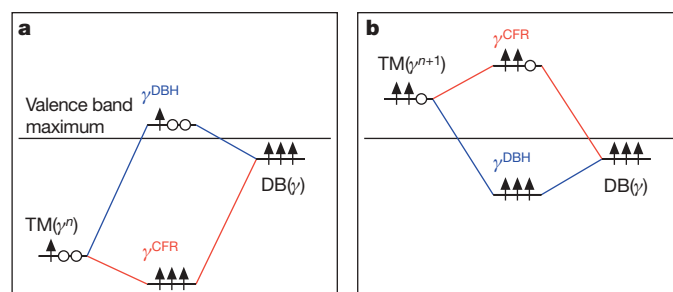


Figure 1 | Energy-level diagrams for TM–host interaction. **a**, System charge, *q*; **b**, system charge, *q* − 1. In each panel, the free-TM-atom orbital with irreducible representation γ (left-hand side) interacts with a host dangling bond level with the same representation (right-hand side), forming two hybrid levels (centre), namely the crystal field resonance level γ^{CFR} and a dangling bond hybrid level γ^{DBH} . The arrows indicate spin-up electrons occupying the level, and the circles indicate unoccupied states.

¹National Renewable Energy Laboratory, Golden, Colorado 80401, USA.

Note that although the electronic charge associated with the CFR level is localized more on the TM site than on the ligands, and that of the DBH level is localized more on the ligands than on the TM site, both levels are hybrid states whose relative weight changes upon adding or removing electrons, owing to the negative charge feedback mechanism described in this work.

To quantify the TM response to changes in global charge, we calculate *ab initio* the change in total energy when the occupation of the antibonding gap level is changed, corresponding to a change in system charge from q to q' . These transition energies (donor or acceptor 'levels') are calculated, with respect to the valence band maximum ε_v , as

$$\varepsilon(q/q') = \frac{E(q') - E(q)}{q - q'} - \varepsilon_v$$

where $E(q)$ is the total energy of the system in charge state q , and by convention the slash in the argument of $\varepsilon(q/q')$ serves to indicate transition between charge states (not division of the charges). These transition energies are given in Figs 2 and 3. We see that all of the considered 3d impurities have such transition levels in the host crystal band gap, consistent also with experiment¹⁸. The effective Mott–Hubbard U

$$U = E(q + 1) + E(q - 1) - 2E(q) = \varepsilon(q - 1/q) - \varepsilon(q/q + 1) \quad (1)$$

is defined as the difference between two successive 'transition energies', for example $\varepsilon(0/-1) - \varepsilon(+1/0)$.

Figures 2 and 3 show that for charge states q that are stable in the gap, U ranges roughly from 0.4 to 3.3 eV. This 'system U ' characterizing the combined host–impurity system is one to two orders smaller

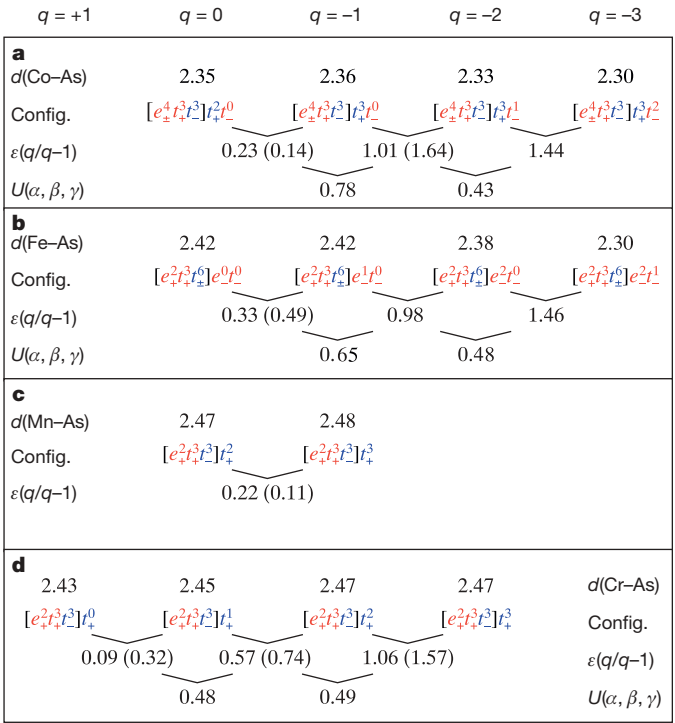


Figure 2 | Properties of GaAs:TM systems, as functions of stable system charge q . **a**, GaAs:Co; **b**, GaAs:Fe; **c**, GaAs:Mn; **d**, GaAs:Cr. In each panel we show bond lengths $d(\text{TM}–\text{As})$ (Å; first line), single-particle level configurations (second line), transition energies $\varepsilon(q/q - 1)$ (eV; third line) and Coulomb energies $U(\alpha, \beta, \gamma)$ (eV; fourth line). The single-particle levels resonant in the valence band are shown in square brackets, CFRs are shown in red and DBHs are shown in blue. The transition energies are given for successive charge states q and $q - 1$, and the Coulomb energies $U(\alpha, \beta, \gamma)$ corresponding to single-particle level configurations α, β and γ for successive transition levels. Experimental values for the transition levels¹⁸ are given in parentheses.

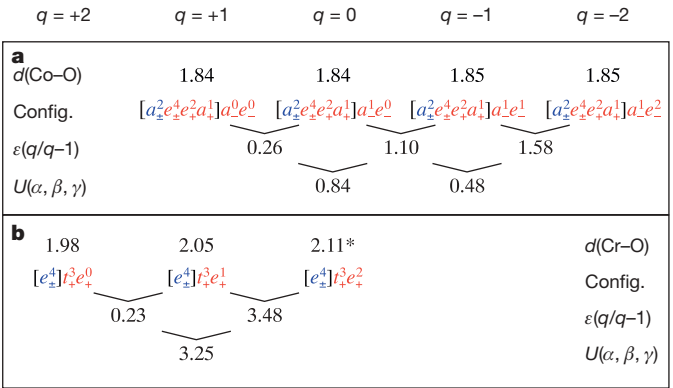


Figure 3 | Properties of $\text{Cu}_2\text{O}:\text{Co}$ and $\text{MgO}:\text{Cr}$, as functions of stable system charge q . **a**, $\text{Cu}_2\text{O}:\text{Co}$; **b**, $\text{MgO}:\text{Cr}$. As in Fig. 2, with bond lengths $d(\text{Co}–\text{O})$ (**a**) and $d(\text{Cr}–\text{O})$ (**b**). We note that the charge-neutral state for $\text{MgO}:\text{Cr}$ (denoted with an asterisk) is Jahn–Teller distorted: four O ions have a Cr–O distance of 2.11 Å, whereas two O ions relax outwards and have a Cr–O distance of 2.29 Å.

than the Mott–Hubbard U for the free TM atom, which is $U_{\text{atom}} \approx 30$ eV (Table 1). The existence of multiple charge configurations within a small semiconductor band gap with this large reduction in U was explained by Haldane and Anderson¹⁹ by changes in hybridization. From the point of view of polarization theory, U_d , which describes a TM(d) level embedded in an insulating system, was empirically found⁶ to be approximately described by the relation $U_d = U_{\text{atom}}(1 - 1/\varepsilon_{\text{eff}})e^2/R$, where e is the electron charge, ε_{eff} is an effective constant for dielectric screening and R is the ionic radius (tabulated in ref. 5). Such polarization and screening effects on U are naturally included in our density-functional calculations, but we emphasize here the importance of the non-rigid electronic level transformation (Fig. 1). This usually receives little attention in ionic systems, where the change of gap-level occupation (increase in Q_{gap} ; see charge definitions below) is often interpreted as a literal (rigid-band-like) TM ionization (increase in Q_{tot})^{3,4,13–16}.

Although there is no unambiguous way to determine an ionic charge in a solid, the ionization of an atom (that is, a literal charge transfer) should be reflected by a change in the integrated electronic charge within a fixed region of space around the atomic site. We calculate the charge Q_{tot} due to all occupied levels below the Fermi energy inside a sphere of radius R centred at the atomic position of the TM for system charge q , as¹

$$Q_{\text{tot}}(q) = \int_0^R dr \rho(r) = \sum_i^{\text{occ.}} \int_0^R dr \psi_i^2$$

where the sum is taken over all occupied levels, ψ_i is the wavefunction projected on TM(d) orbitals, ρ is the charge density and R is chosen as 1.3 Å. The total charge Q_{tot} is further decomposed into the charge Q_{gap} corresponding to the antibonding gap level (summing over all occupied gap levels) and the charge Q_{VB} corresponding to the bonding TM states (summing over all levels below the valence band maximum) resonant within the valence band ($Q_{\text{tot}} = Q_{\text{gap}} + Q_{\text{VB}}$):

Table 1 | Mott–Hubbard U (eV) for the free ions, calculated as defined in equation (1)

TM	$U(d^5, d^4, d^3)$	$U(d^4, d^3, d^2)$	$U(d^3, d^2, d^1)$
Co	27.3	29.5	32.9
Fe	26.8	27.1	27.5
Mn	28.0	23.8	29.3
Cr	25.2	29.5	22.3

The free-atom total energies are calculated using the same GGA–PBE functional as used in the host–impurity calculations.

$$Q_{\text{gap}}(q) = \sum_i^{\text{gap}} \int_0^R dr \psi_i^2$$

$$Q_{\text{VB}}(q) = \sum_i^{\text{VB}} \int_0^R dr \psi_i^2$$

Figure 4 shows the integrated charge quantities Q_{tot} , Q_{gap} and Q_{VB} as functions of electronic configuration, or system charge q . Increasing the system charge by populating the antibonding gap levels (Fig. 1) causes the corresponding gap-level charge Q_{gap} around the TM site to increase by $0.2e-1e$ per one electron increase in q . In a naive rigid-band picture the total charge Q_{tot} around the TM is expected to increase by the same amount as Q_{gap} increases. However, owing to the negative feedback (in the self-consistent calculation), the charge Q_{VB} around the TM, contributed by the bonding levels resonant in the valence band (Fig. 1), decreases by the amount Q_{gap} increases, keeping the total charge Q_{tot} around the TM almost unchanged. Consequently, for all $3d$ TMs studied in GaAs, Cu_2O and MgO , the total charge inside the TM-centred sphere changes by no more than $0.1e$ when the gap level is occupied or vacated by one electron.

The negative feedback causing the decrease in the charge Q_{VB} upon an increase of Q_{gap} is illustrated in terms of the energy-level shifts shown in Fig. 1a, b: when charge is added to the system, the TM(γ) free-atom levels, together with the corresponding bonding and antibonding levels of the combined system, are shifted to higher energies. Thus, the bonding level becomes more localized on the anion dangling bond, yielding a smaller contribution to the charge around the TM site. The orbital energy levels TM(γ) shift so as to balance out any increase in the charge of the antibonding gap level by 'leaking' charge from the bonding levels to the neighbouring anion sites (which may further leak charge to more distant neighbours). To illustrate this charge regulation, we show in Fig. 5 the spatial redistribution of charge density. We plot the difference in charge density ρ corresponding

to the addition of one electron to the charge-neutral state of Mn in GaAs ($\Delta\rho = \rho(0) - \rho(-1)$; Fig. 5a), and to the charge-+1 states of Co in Cu_2O and Cr in MgO ($\Delta\rho = \rho(+1) - \rho(0)$; Fig. 5b, c). In all these cases, we find a decrease (shown in red) and an increase (shown in blue) in charge density around the TM site that tend to cancel each other out. The spatial shapes of the charge density differences correspond to the expected spatial shapes of the gap level symmetries (t , a and e for GaAs:Mn, Cu_2O :Co and MgO :Cr, respectively), and the increase and decrease have the spatial shape of antibonding and bonding levels, respectively. Thus, the increase in Q_{gap} and concomitant decrease in Q_{VB} shown in Fig. 4 can indeed be identified as an increase in antibonding and a decrease in bonding level charges, as suggested by the model in Fig. 1.

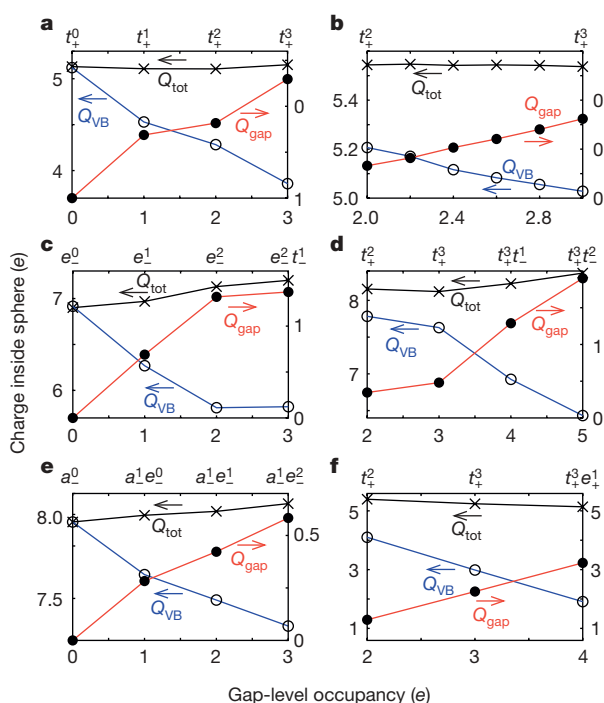


Figure 4 | Integrated charge inside the TM-centred sphere. **a**, GaAs:Cr; **b**, GaAs:Mn; **c**, GaAs:Fe; **d**, GaAs:Co; **e**, Cu_2O :Co; **f**, MgO :Cr. The total charge Q_{tot} is decomposed into contributions from gap-level charge Q_{gap} (red) and from TM charge inside the valence band Q_{VB} (blue).

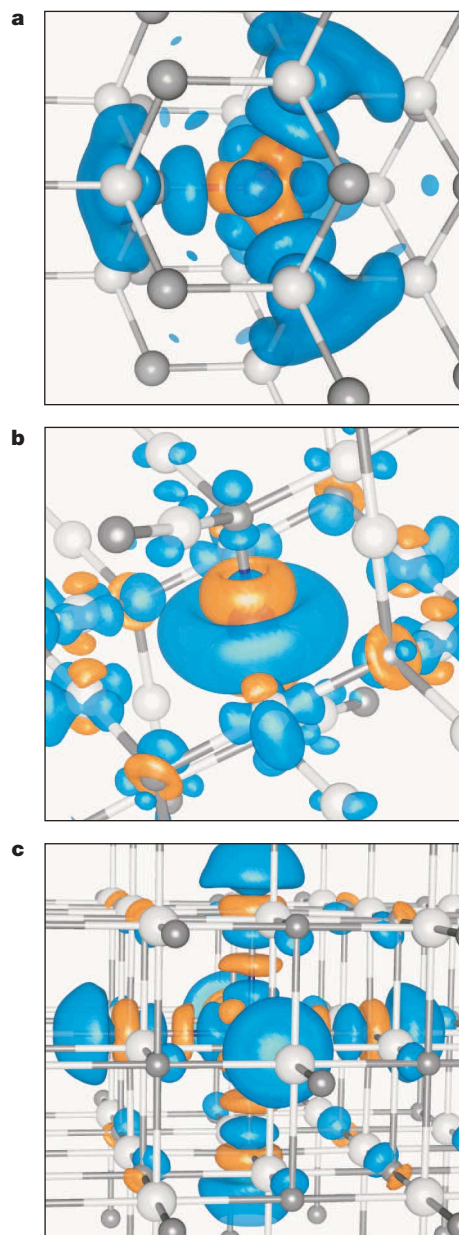


Figure 5 | Charge density differences. **a**, Difference between charge-1 and charge-neutral states for GaAs:Mn; **b**, difference between charge-neutral and charge-+1 states for Cu_2O :Co; **c**, difference between charge-neutral and charge-+1 states for MgO :Cr. Cations and anions are shown as white and grey spheres, respectively. The blue and red isosurfaces respectively correspond to charge densities of $0.003e \text{ \AA}^{-3}$ and $-0.003e \text{ \AA}^{-3}$ (**a**), $0.01e \text{ \AA}^{-3}$ and $-0.01e \text{ \AA}^{-3}$ (**b**), and $0.015e \text{ \AA}^{-3}$ and $-0.015e \text{ \AA}^{-3}$. We note that in each system, upon adding an electron, the blue and red isosurfaces respectively represent an increase and a decrease in the total charge density.

Even though there is no significant charge accumulation around the TM, different oxidation states may be assigned to the host-impurity system on the basis of the type of level being occupied. A change in gap-level occupation is accompanied by a change in the TM oxidation state only when the gap level is a CFR level (Fig. 1b), and not a DBH level (Fig. 1a). Similarly, a change in occupancy of a CFR level affects the local magnetization of the TM, whereas a change in occupancy of a DBH level affects the magnetization of host states instead. Furthermore, the change in spatial distribution of the charge density around the TM site upon changes in gap level occupancy (Fig. 5) influences all TM levels (even the presumably free-atom-like core levels), and can be observed as a core-level shift in an X-ray absorption experiment. (See also Supplementary Information section B.)

The changes in level occupation further affect the lattice relaxations, which for TM ions in oxides and sulphides were traditionally interpreted as ionic radii⁵. These ionic radii are based on the simple assumption that adding electrons to the TM ion increases its radius and removing electrons reduces its radius (order of magnitude of the change in TM ionic radius in an oxide host is typically 0.1 Å per electron added to system⁵). We find, however, by direct total-energy minimization of the TM–anion bond length when electrons are added to the MgO:Cr system (Fig. 3), an outward relaxation of nearest-neighbour O atoms of the same magnitude as reported in ref. 5, even when the local charge around the Cr atom does not increase (Fig. 4). Our results show that this increase in ionic radii upon reduction (adding electrons) simply reflects an increased occupation of the antibonding gap levels instead of a charge build-up at the TM atom itself. The magnitude of the inward/outward relaxation reflects the localization of the gap level; that is, when a level is strongly localized around the TM site, as it is for Cr in MgO, the associated local lattice relaxations are very large, whereas for less localized states, like the TM in GaAs or Co in Cu₂O, the lattice relaxations are smaller (Figs 2 and 3), because the antibonding level the occupation of which increases is more delocalized on extended host states.

The negative-feedback charge regulation is inherent to TMs, and thus occurs in virtually any TM coordination compound. Other than TM impurities in semiconductors^{7,8}, further examples include CaMnO₃¹², where Mn occupies two sites that are inequivalent in geometry but equivalent in charge, and the rehybridization of Co in Li_xCoO₂ upon removal of Li (ref. 9), or in Na_xCoO₂ upon removal of Na (ref. 20). This TM rehybridization associated with multiple oxidation states of the TM atom has been observed in manganites, nickelates, vanadates and so on, and is often falsely attributed to ‘charge ordering’^{3,13,14,16}, that is, a model based on explicit charge transfer to or from the TM site. The self-regulating response, however, ensures that the net charge transfer Δq to or from the TM site is negligibly small ($|\Delta q| \ll 1$), contradicting any model based on assumptions of integer changes in ionization or ‘charge disproportionation’. Instead, ‘ordering’ and ‘disproportionation’ in terms of different TM oxidation states (reflecting different orbital occupations) occurs at essentially constant TM charge, which is a consequence of the present negative charge-feedback mechanism.

METHODS SUMMARY

The TM atom embedded in a crystalline semiconductor or ionic host material was studied in supercells of 64–128 atoms with one host cation replaced with a 3d transition element. The global charge state was altered by occupying the energy levels appearing in the band gap by a different number of electrons, and a

universal compensating background charge was added to maintain overall neutrality. Total energies and band structures were calculated within the generalized gradient approximation (GGA-PBE) to the density-functional formalism, as implemented in the VASP package^{21,22}. Total energies and eigenvalues were corrected for potential alignment and image charges as described in the appendixes of ref. 23.

Received 15 December 2007; accepted 14 April 2008.

1. Wells, A. F. *Structural Inorganic Chemistry* (Clarendon, Oxford, UK, 1975).
2. Cotton, F. A. & Wilkinson, G. *Advanced Inorganic Chemistry* (Wiley, New York, 1988).
3. Goodenough, J. B. & Rivadulla, F. Bond-length fluctuations in transition-metal oxides. *Mod. Phys. Lett. B* **22**, 1057–1081 (2005).
4. Solomon, E. I., Hedman, B., Hodgson, K. O., Dey, A. & Szilagy, R. K. Ligand K-edge X-ray absorption spectroscopy: covalency of ligand-metal bonds. *Coord. Chem. Rev.* **249**, 97–129 (2005).
5. Shannon, R. D. & Prewitt, C. T. Effective ionic radii in oxides and fluorides. *Acta Crystallogr. B* **25**, 925–946 (1969).
6. Koster, G., Geballe, T. H. & Mozyzhes, B. Charge instabilities in the ionic model of metal oxides: importance of polarization energy. *Phys. Rev. B* **66**, 085109 (2002).
7. Zunger, A. & Lindefelt, U. Substitutional 3d impurities in silicon: a self-regulating system. *Solid State Commun.* **45**, 343–346 (1983).
8. Zunger, A. in *Solid State Physics* Vol. 39 (eds Seitz, F., Turnbull, D. & Ehrenreich, H.) 275–464 (Academic, New York, 1986).
9. Wolverton, C. & Zunger, A. First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li_xCoO₂. *Phys. Rev. Lett.* **81**, 606–609 (1998).
10. Leonov, I., Yaresko, A. N., Antonov, V. N., Korotin, M. A. & Anisimov, V. I. Charge and orbital order in Fe₃O₄. *Phys. Rev. Lett.* **93**, 146404 (2004).
11. Jeng, H.-T., Guo, G. Y. & Huang, D. J. Charge-orbital ordering and Verwey transition in magnetite. *Phys. Rev. Lett.* **93**, 156403 (2004).
12. Luo, W. *et al.* Orbital-occupancy versus charge ordering and the strength of electron correlations in electron-doped CaMnO₃. *Phys. Rev. Lett.* **99**, 036402 (2007).
13. Ikeda, N. *et al.* Ferroelectricity from iron valence ordering in the charge-frustrated system LuFe₂O₄. *Nature* **436**, 1136–1138 (2005).
14. Angst, M. *et al.* Charge order with integer iron valence in Fe₂OBO₃. *Phys. Rev. Lett.* **99**, 086403 (2007).
15. Shim, J. H. & Lee, S. Coexistence of two different Cr ions by self-doping in half-metallic CrO₂ nanorods. *Phys. Rev. Lett.* **99**, 057209 (2007).
16. Mazin, I. I. *et al.* Charge ordering as alternative to Jahn-Teller distortion. *Phys. Rev. Lett.* **98**, 176406 (2007).
17. Mahadevan, P., Zunger, A. & Sarma, D. D. Unusual directional dependence of exchange energies in GaAs diluted magnetic semiconductors with Mn: Is the RKKY description relevant? *Phys. Rev. Lett.* **93**, 177201 (2004).
18. Clerfaut, B. Transition-metal impurities in III–V compounds. *J. Phys. C* **18**, 3615–3661 (1985).
19. Haldane, F. D. M. & Anderson, P. W. Simple model of multiple charge states of transition-metal impurities in semiconductors. *Phys. Rev. B* **13**, 2553–2559 (1976).
20. Karppinen, M., Asako, I., Motohashi, T. & Yamauchi, H. Oxygen nonstoichiometry and actual Co valence in Na_xCoO_{2-δ}. *Phys. Rev. B* **71**, 092105 (2005).
21. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
22. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
23. Persson, C., Zhao, Y. J., Lany, S. & Zunger, A. *n*-type doping in CuInSe₂ and CuGaSe₂. *Phys. Rev. B* **72**, 035211 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H.R. thanks G. Trimarchi and J. Chan for discussions and for reading the manuscript. A.Z. thanks P. Mahadevan for interest in the early stages of this problem. This work was funded by the US Department of Energy, Office of Science, under NREL Contract No. DE-AC36-99G010337.

Author Contributions H.R. carried out the calculations, analysed the results and wrote the paper. S.L. and A.Z. contributed to the design of the study, the analysis of results and the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.R. (hannes_raebiger@nrel.gov) or A.Z. (alex_zunger@nrel.gov).

Hydrogen sulphide release to surface waters at the Precambrian/Cambrian boundary

Martin Wille¹, Thomas F. Nägler¹, Bernd Lehmann², Stefan Schröder^{3†} & Jan D. Kramers¹

Animal-like multicellular fossils appeared towards the end of the Precambrian, followed by a rapid increase in the abundance and diversity of fossils during the Early Cambrian period, an event also known as the 'Cambrian explosion'^{1–3}. Changes in the environmental conditions at the Precambrian/Cambrian transition (about 542 Myr ago) have been suggested as a possible explanation for this event, but are still a matter of debate^{1–3}. Here we report molybdenum isotope signatures of black shales from two stratigraphically correlated sample sets with a depositional age of around 542 Myr. We find a transient molybdenum isotope signal immediately after the Precambrian/Cambrian transition. Using a box model of the oceanic molybdenum cycle, we find that intense upwelling of hydrogen sulphide-rich deep ocean water best explains the observed Early Cambrian molybdenum isotope signal. Our findings suggest that the Early Cambrian animal radiation may have been triggered by a major change in ocean circulation, terminating a long period during which the Proterozoic ocean was stratified, with sulphidic deep water.

Trace-metal records, especially the molybdenum (Mo) isotopic record, are increasingly used as a proxy for changing redox conditions in the oceans^{4–6}. Mo is highly soluble and unreactive under oxic conditions, but is easily incorporated into sediments under highly euxinic conditions. Specifically, the Mo isotopic seawater record as deduced mainly from euxinic sediments reflects the relative global amount of euxinic Mo sedimentation and, therefore, the redox condition of the oceanic environment. Thus, Mo signatures in black shales of the Precambrian/Cambrian (PC/C) transition can provide insight into the redox condition of the ocean during that period. Here we present and compare Mo isotopic signatures from black shales of the Ara Group, Oman^{7,8}, and from the Niutitang Formation, Yangtze Platform, China⁹, and examine their global significance (Fig. 1). The Ara Group contains an ash bed within the basal strata of the A4 carbonate bed, dated to 542.0 ± 0.3 Myr ago³, which is correlative to the base of the black shales studied here. The Chinese black shale sequence of the Niutitang Formation has a basal sulphide marker bed with an Re–Os age of 541 ± 16 Myr (ref. 10). The correlation of the two profiles is further corroborated by the globally recorded negative carbon isotopic excursion at the PC/C boundary^{3,11}.

Samples from the Chinese sulphide marker bed from different localities, with extreme Mo concentrations up to 7 wt%, show a homogeneous Mo isotopic signal of around $1.1\delta^{98/95}\text{Mo}$ (where $\delta^{98/95}\text{Mo}$ denotes the parts-per-thousand deviation of the $^{98}\text{Mo}/^{95}\text{Mo}$ ratio relative to the J&M standard). The same overall Mo isotopic signal is also found in black shales stratigraphically above the sulphide marker bed at much lower Mo concentrations. Although hydrothermal circulation has been proposed as a source for metal enrichment¹², the homogeneous isotopic composition over a large Mo concentration range can only originate from a single, isotopically

homogeneous Mo source⁹. A similar overall Mo isotopic composition of $1.2\delta^{98/95}\text{Mo}_{\text{J&M standard}}$ is found in the Early Cambrian black shales from Oman, lending support to the interpretation of sea water being the main source of the Mo signal in both basins.

Mid-Proterozoic-era Mo isotopic signatures show similar values and are interpreted as evidence for widespread anoxia in the deep mid-Proterozoic oceans⁶. Our new Early Cambrian data support the idea that this ocean redox stratification was long-lasting. If the Mo source of the metal-enriched sulphide layer at the PC/C boundary is the Mo sea water inventory⁹, then the Mo-scavenging mechanism involved cannot be the same as that responsible for metal enrichment in recent euxinic sediments, where increased total organic carbon supply is considered as the main driving mechanism¹³. The total organic carbon concentrations within the Chinese sulphide marker bed are similar to those of normal black shales (~ 10 wt%), whereas the Mo concentrations are several orders of magnitude higher.

The transformation from molybdate to thiomolybdate in the presence of H_2S is a highly efficient Mo-scavenging process¹⁴. However, the supply of molybdate to the anoxic water mass is equally important for Mo enrichment¹³. The high Mo concentrations found in the Chinese marker bed require both excessive H_2S concentration and sufficient molybdate supply. The mixing of euxinic, H_2S -rich bottom waters with molybdate-rich surface waters in an upwelling regime

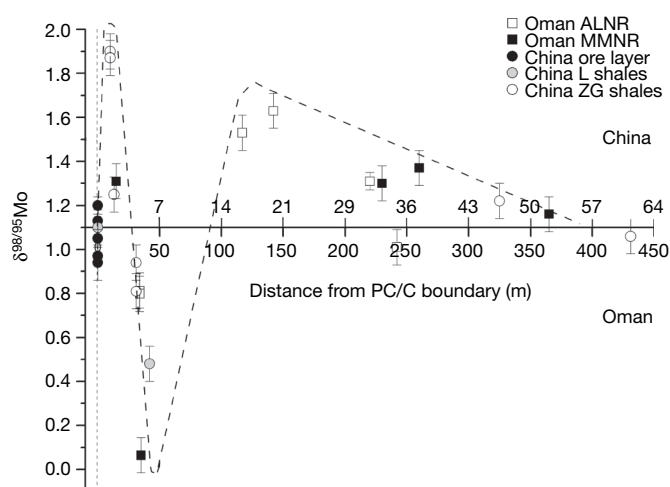


Figure 1 | Merged Mo transient signal of Early Cambrian black shales from the Yangtze Platform and Oman basin. Oman MMNR and Oman ALNR represent black shales from intermediate and deep basins wells, respectively⁷. Chinese ZG and L shales are from the Niutitang Formation (ZG taken at Ganziping, L taken at Yuanling)⁹. Error bars show the external standard reproducibility of $0.1\delta^{98/95}\text{Mo}$ if measured uncertainties are lower.

¹Institute of Geological Sciences, University of Bern, Baltzerstrasse 3, 3012 Bern, Switzerland. ²Institute of Mineralogy and Mineral Resources, Technical University of Clausthal, 38678 Clausthal-Zellerfeld, Germany. ³Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: Total E&P, Avenue Larribau, F-64018 Pau, France.

could thus be a reasonable explanation for the extreme metal enrichment found in the Chinese sulphide marker bed and other localities¹⁵. Additionally, the redox-sensitive elements such as uranium and vanadium also show enrichments in both basins at the PC/C boundary (see Supplementary Information) and provide evidence for euxinic conditions in both basins¹⁶.

Further evidence for such an explanation comes from Mo isotopic signatures of black shales immediately above the PC/C boundary, which in both basins show the largest offset from the overall homogeneous Mo isotopic composition of $\sim 1.1\delta^{98/95}\text{Mo}$. A quasi-linear trend from heavier to lighter Mo isotopic composition can be identified within the first few metres in both basins. The slopes of the two regression lines through these data differ by a factor of seven. Assuming a common environmental process to be responsible for the decline observed in both basins, a sedimentation rate that is seven times higher in Oman than in South China would be the most straightforward explanation. If the complete sections are correspondingly scaled by this factor, a covariance of Mo isotopes is also found for the remaining stratigraphy (Fig. 1). This assumption is supported by the decrease in uranium and vanadium abundances stratigraphically upwards from the PC/C boundary, which indicates a sedimentation rate for Oman^{7,16} that is five to ten times higher than that for South China (see Supplementary Information for details). The scaled and combined Mo isotopic signal (Fig. 1) indicates a transient Mo signal immediately following the PC/C boundary, most likely resulting from a brief, global, non-steady state situation within the earliest Cambrian ocean.

The palaeoenvironmental implications of this Early Cambrian transient Mo isotope signal can be evaluated by modelling the Mo oceanic inventory in relation to the Mo fluxes. The Mo seawater composition can be adequately estimated by adjusting the Mo sedimentary fluxes in a first-order box model of the Mo oceanic cycle. We divided the Mo output into variable proportions of a strong euxinic output (reflecting the seawater Mo isotopic composition) and a light output (with a Mo signature that is $-2\delta^{98/95}\text{Mo}$ lighter than sea water). The latter model parameter represents the sum of oxic, suboxic

and even slightly euxinic Mo sedimentations, all of which are known to incorporate Mo isotopic compositions lower than that of coexisting sea water^{17,18}. Which of these light sinks is dominant is to be deduced from the geological context (see below). Under steady-state conditions the oceanic Mo input (oxic continental weathering) must equal its output. We have also assumed Mo input values and the Mo abundance in sea water to be the same as today's values. Changing these parameters would result in a change of absolute model timescales. However, relative timescales and processes would remain unaffected.

The necessary conditions to mimic the transient Mo signal are shown in two model runs in Fig. 2. In both scenarios, the Mo ocean inventory is depleted on a very short timescale owing to a euxinic Mo output increased by a factor of more than 150 relative to the previous steady state. This strong depletion of the Mo ocean inventory is necessary to make the Mo isotopic composition of the continental Mo influx ($\sim 0.2\delta^{98/95}\text{Mo}$) the dominant signature for a short time as observed in the large negative $\delta^{98/95}\text{Mo}$ peak shortly after the PC/C boundary. Another requirement is the short timescale of this euxinic output. Both model results show that the sudden increase in the euxinic output is followed by an exponential decline, which returns the Mo euxinic output to steady-state values within around 250,000 yr. A disruption of the Mo oceanic steady-state situation as a result of this short, but intense, Mo euxinic output accounts for the fluctuation of the Mo signal on such short timescales.

To reconcile the two positive Mo isotopic excursions seen in the Mo isotopic data set, an additional 30-times-increased light Mo output has to be implemented in the model (Fig. 2b). The combination of the increased Mo euxinic output together with the increased light/suboxic fractionation results in Mo depletion and a heavy isotopic composition of the ocean inventory shortly after the PC/C boundary. The probable cause for the second positive Mo excursion is a slower decline in the light/suboxic sedimentation in comparison with the euxinic sedimentation.

Our model results for the Mo isotopic transient signal have two main implications consistent with the metal enrichments found at the PC/C boundary. First, the strong increase in the euxinic output

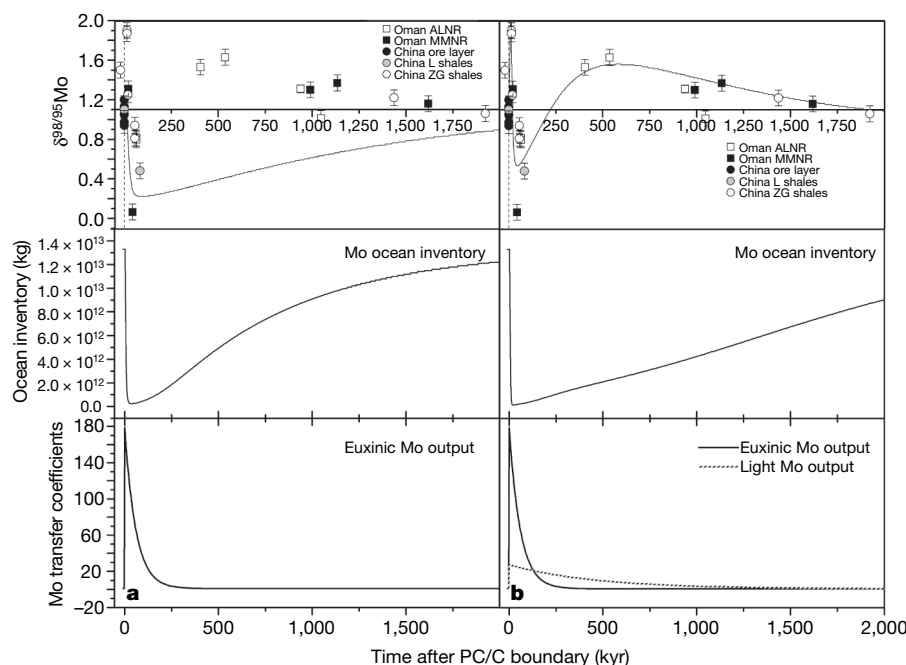


Figure 2 | The modelled Mo isotopic seawater signature in comparison with measured black shale isotope Mo data. Stratigraphic positions of the shale samples were recalculated into deposition time after the PC/C boundary, taking higher authigenic sedimentation rates shortly after the PC/C boundary into account. **a**, The modelled $\delta^{98/95}\text{Mo}$ curve taking only

increased euxinic Mo sedimentation into account. **b**, The modelled $\delta^{98/95}\text{Mo}$ curve with increased euxinic and light Mo output. A light Mo output of $-2\delta^{98/95}\text{Mo}$ was used (see Supplementary Information for details). Error bars show the external standard reproducibility of $0.1\delta^{98/95}\text{Mo}$ if measured uncertainties are lower.

together with its fast exponential decline are in accord with the notion of a short but intense upwelling of euxinic bottom-water masses that would have dramatically lowered the Mo ocean inventory. Variations in the carbon isotopic signal ($\delta^{13}\text{C}$) at the PC/C boundary indicate non-steady-state processes on short timescales ($<100,000\text{ yr}$)¹⁹, which is in good agreement with our modelled upwelling time range. A concurrent increase in weakly euxinic to suboxic sedimentary conditions, manifest in an isotopically light Mo output, can be expected in upwelling regions owing to the mixture of bottom-water H_2S and oxic near-surface waters and the oxidation of H_2S at the water–atmosphere interface.

This picture is consistent with other prominent geological and geochemical features at the PC/C boundary, such as the negative $\delta^{13}\text{C}$ excursion and the global accumulation of large phosphorite deposits²⁰. Organic matter sinking into the anoxic bottom-water zone resulted in the build-up of a dissolved phosphate and isotopically light organic carbon repository²¹. The destabilization of the oceanic chemocline²², possibly caused by a previous increase in oxygen levels²³ or by a change in ocean circulation patterns, could have brought these anoxic bottom-water masses to shelf regions containing free oxygen. Oxidation would have released phosphorus, which was then predominantly mineralized²⁴. Subsequently, the isotopically light carbon supply into shelf regions was probably responsible for the prominent negative $\delta^{13}\text{C}$ excursion.

The fact that most Ediacaran fossils have no post-Proterozoic record and that Early Cambrian trace and body fossils appear over only a protracted interval in the Cambrian suggests that the Cambrian fauna did not simply succeed the Ediacaran biota²⁵. This time gap lends substance to the hypothesis that the Ediacaran and Cambrian faunas are separated by an event of mass extinction. Several palaeoclimatic, palaeoenvironmental and palaeoecological reasons can be responsible for mass extinctions in general²⁶. These include catastrophic methane release²⁷, large-scale volcanism (both of which would lead to global warming and hypercapnia), and hydrogen sulphide poisoning due to upwelling euxinic bottom waters. Of these, hydrogen sulphide poisoning is the only process that can account for the observed Early Cambrian Mo isotope signal and provide a plausible explanation for the sudden extinction of the Ediacaran fauna. It is well established that hydrogen sulphide is almost universally toxic to eukaryotic cells in micromolar or higher concentrations^{22,28}. Increasing oxygenation of the upper ocean and atmosphere following the Marinoan glaciation is thought to have triggered the development of the Ediacaran fauna²³, and to have caused destabilization of the oceanic chemocline²³. As a consequence, upwelling euxinic bottom water is likely to have rapidly poisoned the Ediacaran fauna.

METHODS SUMMARY

Sample powders were oxidized at 800°C for $\sim 8\text{ h}$. The equivalent of $\geq 50\text{ ng Mo}$ was put in a Teflon beaker together with 6 M HCl and heated to $\sim 100^\circ\text{C}$ for $\sim 24\text{ h}$. The supernatant was decanted into a second beaker containing the necessary amount of ^{97}Mo – ^{100}Mo double spike. The remaining sample material was dried and attacked with concentrated HF and HNO_3 (4:1) at $\sim 100^\circ\text{C}$. After complete dissolution and evaporation, the residue was redissolved in 6 M HCl and added to the second beaker. After evaporation, the material was taken up in $4\text{ M HCl} + \text{H}_2\text{O}_2$ and loaded onto an anion exchange resin to wash out cations. The molybdate anion was finally eluted with 2 M HNO_3 . A subsequent cation exchange column separated Mo from residual Fe and Zr. Samples were measured as 0.5 M HNO_3 solution on a Nu Instruments MC-ICP-MS equipped with an ESI Apex nebulizer. The external standard reproducibility was $0.1\delta^{98/95}\text{Mo}$ (2 s.d.)⁴ (see Supplementary Information).

Received 18 July 2007; accepted 9 May 2008.

Published online 28 May 2008.

- Brasier, M. Background to the Cambrian Explosion. *J. Geol. Soc. Lond.* **149**, 585–587 (1992).
- Grotzinger, J., Bowring, S., Saylor, B. & Kaufman, A. Biostratigraphic and geochronologic constraints on early animal evolution. *Science* **270**, 598–604 (1995).

- Amthor, J. *et al.* Extinction of Cloudina and Namacalathus at the Precambrian–Cambrian boundary in Oman. *Geology* **31**, 431–434 (2003).
- Siebert, C., Nagler, T., von Blanckenburg, F. & Kramers, J. Molybdenum isotope records as a potential new proxy for paleoceanography. *Earth Planet. Sci. Lett.* **211**, 159–171 (2003).
- Siebert, C., Kramers, J., Meisel, P. & Nagler, T. PGE, Re–Os and Mo isotope systematics in Archean and early Proterozoic sedimentary systems as proxies for redox conditions of the early Earth. *Geochim. Cosmochim. Acta* **69**, 1787–1801 (2005).
- Arnold, G., Anbar, A., Barling, J. & Lyons, T. Molybdenum isotope evidence for widespread anoxia in mid-Proterozoic oceans. *Science* **304**, 87–90 (2004).
- Schroder, S. & Grotzinger, J. Evidence for anoxia at the Ediacaran–Cambrian boundary: The record of redox sensitive trace elements and rare earth elements in Oman. *J. Geol. Soc. Lond.* **164**, 175–187 (2007).
- Amthor, J., Ramseier, K., Faulkner, T. & Lucas, P. Stratigraphy and sedimentology of a chert reservoir at the Precambrian–Cambrian Boundary: the Al Shomou silicilyte, South Oman Salt Basin. *GeoArabia* **10**, 89–122 (2005).
- Lehmann, B. *et al.* Highly metalliferous carbonaceous shale and Early Cambrian seawater. *Geology* **35**, 403–406 (2007).
- Mao, J. *et al.* Re–Os dating of polymetallic Ni–Mo–PGE–Au mineralization in Lower Cambrian black shales of South China and its geological significance. *Econ. Geol.* **97**, 1051–1061 (2002).
- Cao, S., Ma, D. & Pan, J. Stable isotopic geochemistry of organic carbon and pyrite sulfur from the Early Cambrian black shales in Northwestern Hunan China. *Prog. Nat. Sci.* **14**, 181–187 (2004).
- Steiner, M., Wallis, E., Erdtmann, B.-D., Zhao, Y. & Yang, R. Submarine-hydrothermal exhalative ore layer in black shales from South China and associated fossils – insights into Lower Cambrian facies and bio-evolution. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **169**, 165–191 (2001).
- Algeo, J. & Lyons, T. Mo–total organic carbon covariation in modern anoxic marine environments: Implications for analysis of paleoredox and paleohydrographic conditions. *Paleoceanography* **21**, 1–23 (2006).
- Erickson, B. & Helz, G. Molybdenum(VI) speciation in sulfidic waters: Stability and lability of thiomolybdates. *Geochim. Cosmochim. Acta* **64**, 1149–1158 (2000).
- Banerjee, D., Schidlowski, M., Siebert, F. & Brasier, M. Geochemical changes across the Proterozoic–Cambrian transition in the Durmala phosphorite mine section, Mussoorie Hills, Garhwal Himalaya, India. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **132**, 183–194 (1997).
- Pan, J., Ma, D. & Cao, S. Trace element geochemistry of the Lower Cambrian black rock series from northwestern Hunan, South China. *Prog. Nat. Sci.* **14**, 64–70 (2004).
- Siebert, C., McManus, J., Bice, A., Poulson, R. & Berelson, W. Molybdenum isotope signatures in continental margin marine sediments. *Earth Planet. Sci. Lett.* **241**, 723–733 (2006).
- Nagler, T., Siebert, C., Luschen, H. & Bottcher, M. Sedimentary Mo isotope record across the Holocene fresh-brackish water transition of the Black Sea. *Chem. Geol.* **219**, 283–295 (2005).
- Maloof, A., Schrag, D., Crowley, J. & Bowring, S. An expanded record of Early Cambrian carbon cycling from the Anti-Atlas Margin, Morocco. *Can. J. Earth Sci.* **42**, 2195–2216 (2005).
- Cowie, J. & Brasier, M. (eds) *The Precambrian–Cambrian Boundary* (Clarendon, Oxford, UK, 1989).
- Shen, Y., Schidlowski, M. & Chu, X. Biogeochemical approach to understanding phosphogenic events of the terminal Proterozoic to Cambrian. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **158**, 99–108 (2000).
- Kump, L., Pavlov, A. & Arthur, M. Massive release of hydrogen sulfide to the surface ocean and atmosphere during intervals of ocean anoxia. *Geology* **33**, 397–400 (2005).
- Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444**, 744–747 (2006).
- Mort, H. *et al.* Phosphorus and the role of productivity and nutrient recycling during oceanic anoxic event 2. *Geology* **35**, 483–486 (2007).
- Knoll, A. & Carroll, S. Early animal evolution: Emerging views from comparative biology and geology. *Science* **284**, 2129–2137 (1999).
- Twitchett, R. The paleoclimatology, paleoecology and paleoenvironmental analysis of mass extinction events. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **131**, 190–213 (2006).
- Kirschvink, J. & Raub, T. A methane fuse for the Cambrian explosion: carbon cycles and true polar wander. *C.R. Geosci.* **335**, 65–78 (2003).
- Knoll, A., Bambach, R., Payne, J., Pruss, S. & Fischer, W. Paleophysiology and end-Permian mass extinction. *Earth Planet. Sci. Lett.* **256**, 295–313 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was financed by grants from the Swiss National Foundation to J.D.K. and T.F.N., from Deutsche Forschungsgemeinschaft to B.L., and from the German Academic Exchange Service to S.S. Thanks to Petroleum Development Oman LLC for financial and logistical support, and to J. Grotzinger and NASA for initial analytical work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.W. (martin.wille@anu.edu.au).

LETTERS

Simultaneous teleseismic and geodetic observations of the stick-slip motion of an Antarctic ice stream

Douglas A. Wiens¹, Sridhar Anandakrishnan², J. Paul Winberry² & Matt A. King³

Long-period seismic sources associated with glacier motion have been recently discovered^{1,2}, and an increase in ice flow over the past decade has been suggested on the basis of secular changes in such measurements³. Their significance, however, remains uncertain, as a relationship to ice flow has not been confirmed by direct observation. Here we combine long-period surface-wave observations with simultaneous Global Positioning System measurements of ice displacement to study the tidally modulated stick-slip motion of the Whillans Ice Stream in West Antarctica^{4,5}. The seismic origin time corresponds to slip nucleation at a region of the bed of the Whillans Ice Stream that is likely stronger than in surrounding regions and, thus, acts like an 'asperity' in traditional fault models. In addition to the initial pulse, two seismic arrivals occurring 10–23 minutes later represent stopping phases as the slip terminates at the ice stream edge and the grounding line. Seismic amplitude and average rupture velocity are correlated with tidal amplitude for the different slip events during the spring-to-neap tidal cycle. Although the total seismic moment calculated from ice rigidity, slip displacement, and rupture area is equivalent to an earthquake of moment magnitude seven (M_w 7), seismic amplitudes are modest (M_s 3.6–4.2), owing to the source duration of 20–30 minutes. Seismic radiation from ice movement is proportional to the derivative of the moment rate function at periods of 25–100 seconds and very long-period radiation is not detected, owing to the source geometry. Long-period seismic waves are thus useful for detecting and studying sudden ice movements but are insensitive to the total amount of slip.

The mass balance of continental ice sheets is in large part controlled by the flow rates of ice streams and outlet glaciers. The dynamics of ice flow, and in particular the influence of basal properties, has received considerable attention in recent years because of concern about possible rapid future rises in sea level due to the acceleration of these ice streams^{6–8}. Numerical models of ice flow include sliding and basal deformation physics; however, they are incapable of explaining the recent observations which suggest that some motion may occur through stick-slip sliding, similar to that which occurs in earthquake faults⁹. Outlet glaciers in Greenland and several other locations produce globally recorded long-period seismic waves¹. It has been suggested that these events result from sudden glacier motion with a timescale of about 50 s (ref. 2), and that the increasing frequency of these events in recent years may be related to climate change³. However, the mechanics and interpretation of these 'glacial earthquakes' remain highly uncertain, as no simultaneous observations have been published from the source regions.

Global Positioning System (GPS) observations show that the ice plain of the Whillans Ice Stream (WIS, formerly Ice Stream B) in West Antarctica undergoes tidally modulated stick-slip motion twice per day^{4,5}. During each slip event, the ice plain of the WIS (which is

greater than 200 km × 100 km in area and approximately 600 m thick) moves by up to 70 cm in approximately 25 min. The expression for the seismic moment of an elastic dislocation¹⁰, $M_0 = \mu DA$, where μ is the rigidity of glacial ice ($3.5 \times 10^9 \text{ N m}^{-1}$), D is the dislocation and A is the slip area, shows that this event is equivalent to a M_w -7.0 earthquake. However, because of the slow average rupture velocity ($\sim 150 \text{ m s}^{-1}$) and long slip time at a given location, its source duration is nearly two orders of magnitude longer than that of a similarly sized tectonic earthquake. In the years 2001–2003, a deployment of 43 broadband seismographs in the Trans-Antarctic Mountains and East Antarctic interior (approximately 1000 km from the WIS) detected Rayleigh waves originating from the WIS that correlated with the expected times of tidally triggered slip¹¹.

Simultaneous seismic and geodetic observations made during 2004 allow us to determine the mechanism for glacial generation of seismic waves and to better constrain the slip characteristics of the WIS. An array of 19 GPS instruments (see Supplementary Table 1) operating at a sampling rate of 0.1 Hz was deployed on the WIS to record the ice stream motion as part of the Tidal Modulation of Ice-Stream Flow (TIDES) project¹². The ice stream slip events were also well recorded at the borehole Global Seismic Network stations near the South Pole (station QSPA; 89.928° S, 145.0° E) and in the McMurdo Dry Valleys (station VNDA; 77.517° S, 161.853° E) at distances of 620 and 990 km from the WIS, respectively (Fig. 1, upper inset).

Examination of the GPS records from the WIS (see Methods) shows that ice stream slips repeatedly originate near the same location on the ice plain (Fig. 1). Rupture propagation is about 500–1,000 m s^{-1} within 20–30 km of the nucleation point, but becomes much slower ($\sim 100 \text{ m s}^{-1}$) in more distant regions of the ice stream. Slip magnitude becomes smaller progressively farther upstream of the nucleation point; however, the slip retains significant amplitude down to the grounding line and onto the Ross Ice Shelf. The slips terminate near the grounding line to the northwest of the nucleation point about 20–30 min after initiation. The slips are tidally modulated; one slip episode occurs around the time of the high ocean tide and a second occurs 5–12 h later, with the intervals converging to 12 h during times of low tidal amplitude (neap tide). In contradiction to initial reports⁵, the amount of slip is proportional to the inter-event time and, thus, to the stress accumulated since the last slip event. Times of the slip episodes are given in Supplementary Table 2.

Seismic Love waves are recorded at VNDA and QSPA, and Rayleigh waves at VNDA; Rayleigh waves are not well recorded at QSPA, owing to the source radiation pattern. Seismogram signals are detectable at periods of 20–150 s and typically show three arrivals over a time span of 12–23 min (Fig. 2b). Because the recorded seismic periods are much shorter than the rupture duration, the signal is highly band limited, and the observed seismic radiation should

¹Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA. ²Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ³School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

correspond to rapid changes in the seismic moment rate. The onset of the first seismic arrival, when corrected for surface-wave propagation between source and receiver, occurs 20–150 s after the first onset of slip at the GPS station closest to the inferred rupture source. Thus, the initial seismic pulse records the rapid increase in moment rate resulting from the initial fast expansion of the rupture.

The seismic waveforms show systematic variations as a function of the tidal cycle, with larger delays between the first and third arrivals for smaller Ross Sea ocean tidal amplitudes (Fig. 3a). The GPS observations

suggest that the average rupture velocity is correlated with the tidal amplitude (Fig. 3b). This suggests that the second and third seismic pulses probably radiate from specific spatial regions within the rupture zone, and that their timing is determined by their distances from the rupture origin and the rupture velocity. Differences in the timing of the second and third arrivals relative to the first indicate that they do not originate from the rupture onset location.

We implement a simple grid-search procedure to find the source location that best fits both the seismic arrival times and the average

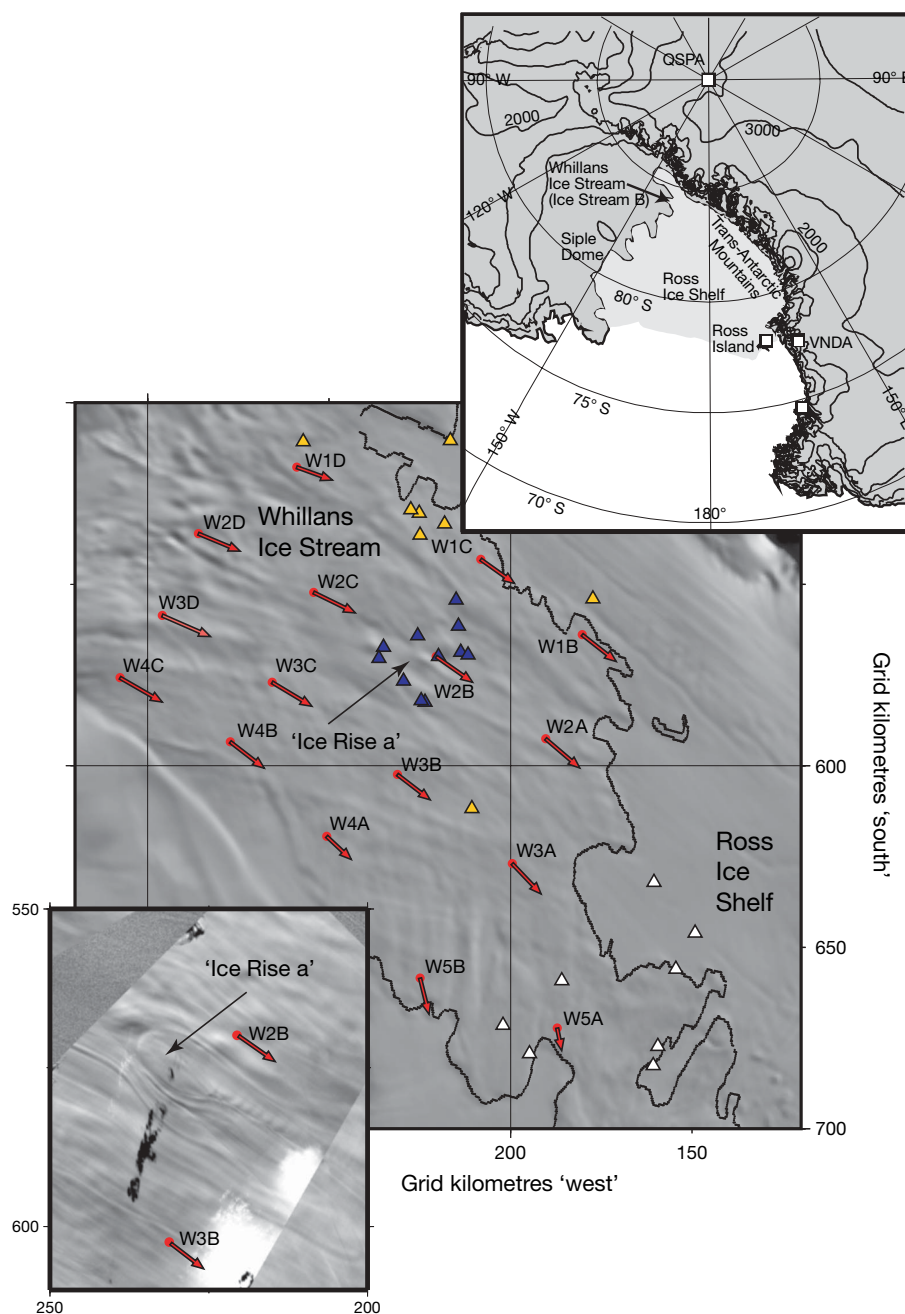


Figure 1 | The Whillans Ice Stream, the positions of the sensors and the positions of the slip nucleation and seismic source locations. TIDES GPS receivers (red circles), the slip nucleation points determined from GPS (blue triangles) and the approximate source regions for the second (orange triangles) and third (white triangles) seismic surface-wave arrivals, calculated from the seismic arrival times and the average rupture velocity for nine slip events from 16–20 November 2004 are shown on a Moderate Resolution Imaging Spectroradiometer mosaic map. The grounding line is marked with a black line²⁴. Red arrows indicate the direction of ice flow determined from GPS, with arrow length proportional to speed (for scale,

the red arrow at W2B represents 392.4 m yr^{-1}). Coordinates are given in a grid system with the origin at the South Pole, the 'south' axis oriented along longitude 180° , and the 'west' axis along longitude 90°W . The upper inset shows the location of the Whillans Ice Stream and the QSPA and VNDA seismic stations, among other permanent seismic stations (open squares). Contours indicate surface elevation in meters. The lower inset, which has the same axes as the main panel, is an enlarged SPOT image showing ice flow features in the region of the rupture nucleation points and 'Ice Rise a', which is thought to represent a region of reduced subglacial water and higher basal friction.

rupture velocity as determined from the GPS observations (see Methods). The source regions of the second seismic arrivals are preferentially located along the southern lateral edge of the ice stream, suggesting that each of the second pulses may result from a rapid decrease in the moment rate when the rupture front reaches that edge. The source regions of the third seismic arrivals are clustered near the grounding line about 100 km northwest of the nucleation point. The origin times calculated from the third seismic arrivals at VNDA and QSPA correspond closely to the arrival times of the peak rupture velocity at GPS receiver W5B (Fig. 1) near the grounding line in this region. This suggests that each third (and final) arrival represents a 'stopping phase' in which the moment rate function begins to decrease owing to the decrease in slip area after the slip reaches the grounding line.

The slip repeatedly nucleates in the same region, near GPS receiver W2B (Fig. 1). The uncertainty in the nucleation point locations is similar to the scatter of the locations, so it is not clear whether or not all the events nucleate at the same point. This site has been previously identified as 'ice rise a' (although the feature is not an ice rise, we continue to refer to it as such)¹³, a patch of higher friction on the ice stream bed inferred on the basis of flow lines and slightly increased ice elevations¹⁴. This region is readily visible in the SPOT (Satellite Pour l'Observation de la Terre) image (Fig. 1, lower inset) and also defines an area of low effective reflection coefficient on radar images¹⁵, which is commonly interpreted as indicating an absence of subglacial water. This suggests that the stick-slip characteristics of the WIS are controlled by an area of higher friction that resembles a classic barrier or

asperity found in earthquake fault models^{16,17}. Thus, a simple explanation consistent with the data is that the WIS motion is affected by both the tidal modulation and by this patch of greater friction, perhaps because of bed composition or hydrological conditions. Resistance at the asperity results in little motion of the ice downstream of the asperity except when loading exceeds the breaking point and a slip event occurs, unlike in the neighbouring Bindschadler Ice Stream, which displays tidally modulated velocities without any slip events¹⁸.

The transverse component of the seismogram at QSPA and the vertical component at VNDA have good signal-to-noise ratios in the 0.012–0.04-Hz frequency band. Although these band-limited data are not sufficient to construct a well-constrained inverse solution, we can obtain significant insight from forward modelling. We model the first packet of surface waves assuming there to be a double-couple source at a shallow depth (1 km); the two later packets are more complicated and variable, probably as a result of the interference of stopping phases from different parts of the slip region.

The modelling shows that, as expected, the seismograms are only sensitive to the spectral components of the moment rate function within the 0.012–0.04-Hz band; the absence of any additional packets

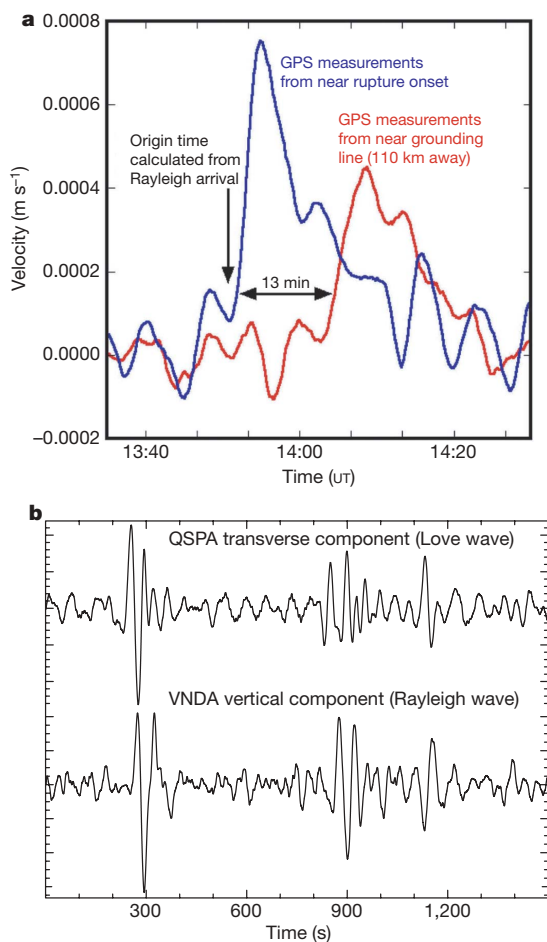


Figure 2 | GPS and seismic records of Whillans slip events. **a**, Ice velocity computed from GPS data recorded near the slip nucleation point (W2B) and near the grounding line (W5A). See Methods for processing details. **b**, Typical seismograms of Whillans slip events recorded at VNDA and QSPA, showing the three main surface-wave packets.

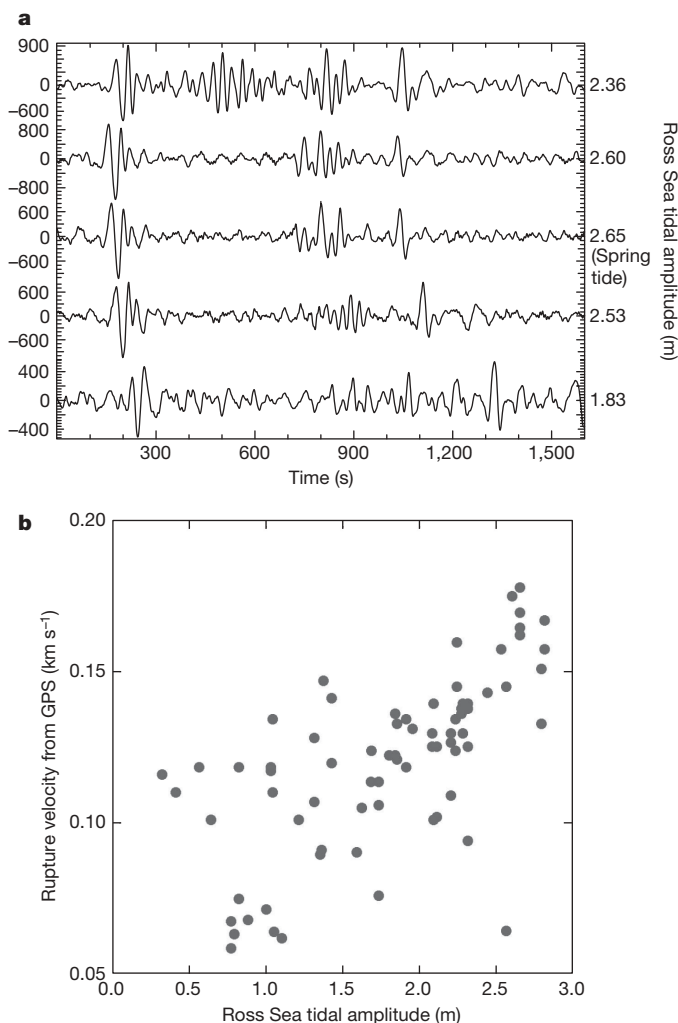


Figure 3 | The effect of Ross Sea tidal amplitude on seismograms and the average rupture velocity from GPS. **a**, Seismograms showing the effects of the tidal cycle on the seismic waveforms (specifically the QSPA transverse component, 25–100 s passband). Rupture velocity correlates with tidal amplitude, resulting in shorter durations for high tides. Tidal amplitude in metres is calculated using model CATS02.01 (ref. 25), and seismogram amplitude is shown in digital counts. **b**, Average rupture velocity determined by GPS as a function of tidal amplitude.

within the first 400 s suggests that the moment rate function is extremely smooth during the rupture expansion phase. In a simple elastic dislocation model, the source time function is the slip area as a function of time convolved with the slip function at any given point along the dislocation. The duration of slip at most GPS sites and the rupture propagation time are both about 700 s, suggesting that the total duration of the time function is approximately 1,400 s. We achieve a good fit to the initial waveforms using an offset inverse cosine function with a period of 800–1,400 s (Fig. 4). The relative amplitudes of the Love and Rayleigh waves at QSPA and VNDA are well fit by a rupture on a plane dipping very shallowly to the west-northwest (strike, 210° ; dip, 3°), consistent with the ice stream motion direction. The observed surface-wave pulse is radiated by the initial part of this function, which has the highest spectral amplitude within the frequency band with good signal.

Because the seismic amplitudes within the passband are controlled by the high-frequency characteristics of the moment rate function rather than its overall amplitude, the total seismic moment of the event, and thus the slip area and amplitude, cannot be constrained by seismic observations alone. The seismic amplitudes are instead a

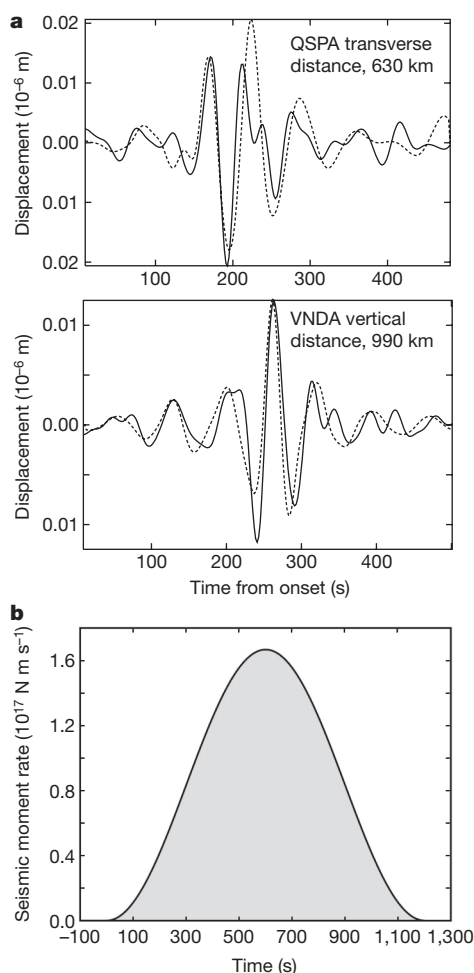


Figure 4 | Modelling the initial seismic signals of the slip events.

a, Observed seismograms (solid lines) of the first Love wave packet (QSPA, upper graph) and Rayleigh wave packet (VNDA, lower graph) for the 0.012–0.04 Hz passband, along with synthetic seismograms (dashed lines) calculated for a dip-slip elastic dislocation dipping 3° towards N 60° W at a depth of 1 km. **b**, Seismic moment rate function used to calculate synthetics in **a**. The total moment release and long-period characteristics of the moment rate function are not well determined, because they are outside the passband of the signal. We note that the later parts of the function involving a decrease in the moment rate are too late to affect the initial waveforms modelled in **a**.

function of the rapidity of the rupture, which is correlated with tidal amplitude, rather than the slip displacement or area. Owing to the extremely long source duration, the seismic amplitudes at periods of 25–100 s are much smaller than would be expected for a tectonic earthquake with this total slip and the assumed elastic properties ($M_w \approx 7.0$). We estimate that the seismic amplitudes correspond to surface wave magnitudes of M_s 3.6 to M_s 4.2, using the method of Russell¹⁹ and extending the period to 50 s. The events are also probably invisible at normal mode periods (200–3,000 s), as several studies searching for undetected sources in this band have failed to find them^{20,21}. This is probably due to the degenerate nature of horizontal dislocations very near the Earth's surface, which causes seismic excitation to go to zero at ultralong periods²², and to the high ambient noise level in this band²³.

The observation of teleseismic surface-wave radiation from the WIS slip events has a number of wider implications. Ekstrom *et al.*¹ discovered long-period seismic sources co-located with glaciers in Greenland and Alaska and at Totten Glacier in Antarctica, using global seismic data, and concluded that they represent glacial slip events. Analysis of these sources using a single-force inversion formalism suggested that the durations of the slip events were on the order of 50 s (ref. 2). There are certainly differences in flow processes between the WIS and the steeper glaciers of the Northern Hemisphere, so it is not yet clear whether our observations are applicable outside the Antarctic. However, the WIS results reported here show that at least for stick-slip motion of large ice streams, the seismic packets represent the onset and termination of much longer (~ 25 min) slip episodes. The amplitudes of these signals are largely a function of the rupture velocity and duration of the slip, rather than the total displacement of the ice. Seismic observations can thus detect which glaciers and ice streams show stick-slip behaviour and identify the times and characteristics of the slip events, but are relatively insensitive to the total amount of ice movement.

METHODS SUMMARY

The GPS antennas were clamped to steel poles driven into the snow surface. The 0.1-Hz dual-frequency GPS data were processed relative to a base station installed on nearby stationary ice. Site motion was loosely constrained at 0.05 m every 10 s, which reduces noise without over-smoothing the signal. The precision of the coordinates was assessed during WIS inter-slip time periods or at linearly moving sites and is ~ 0.01 m and ~ 0.03 m in the horizontal and vertical coordinates, respectively.

One-dimensional positions were calculated from the resulting GPS horizontal-coordinate time series using vector addition. These positions were filtered using a 600-s low-pass filter to reduce temporally correlated noise, and differentiated to produce velocities (Fig. 2a). The locations and times of slip nucleation were found by picking the first point of the filtered time series that rises above the background noise level and solving for the best-fit origin point and rupture velocity using a least-squares grid-search method. The geographical source regions radiating the second and third surface-wave packets were determined using a grid search over the WIS to find the location best fitting the seismic surface-wave arrival times at QSPA and VNDA as well as the average rupture velocity.

We modelled the seismic waveforms assuming that the source can be represented by a double couple at a very shallow depth (~ 1 km). We used a reflectivity algorithm and seismic structures appropriate for West Antarctica and East Antarctica for computing the VNDA and, respectively, QSPA synthetics.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2007; accepted 6 April 2008.

- Ekstrom, G., Nettles, M. & Abers, G. A. Glacial earthquakes. *Science* **302**, 622–624 (2003).
- Tsai, V. C. & Ekstrom, G. Analysis of glacial earthquakes. *J. Geophys. Res.* **112**, doi:10.1029/2006JF000596 (2007).
- Ekstrom, G., Nettles, M. & Tsai, V. C. Seasonality and increasing frequency of Greenland glacial earthquakes. *Science* **311**, 1756–1758 (2006).
- Bindschadler, R. A., King, M. A., Alley, R. B., Anandakrishnan, S. & Padman, L. Tidally controlled stick-slip discharge of a West Antarctic Ice Stream. *Science* **301**, 1087–1089 (2003).

5. Bindschadler, R. A., Vornberger, P. L., King, M. A. & Padman, L. Tidally driven stick-slip motion in the month of Whillans Ice Stream, Antarctica. *Ann. Glaciol.* **36**, 263–272 (2003).
6. Bindschadler, R. A. & Bentley, C. R. On thin ice? *Sci. Am.* **287**, 98–105 (2002).
7. de Angelis, H. & Skvarca, P. Glacier surge after ice shelf collapse. *Science* **299**, 1560–1562 (2003).
8. Scambos, T., Bohlander, J., Shuman, C. & Skvarca, P. Glacier acceleration and thinning after ice shelf collapse in the Larsen B embayment, Antarctica. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020670 (2004).
9. Fischer, U. H. & Clarke, G. K. C. Stick-slip sliding behaviour at the base of a glacier. *Ann. Glaciol.* **24**, 390–396 (1997).
10. Aki, K. Generation and propagation of G waves from the Niigata earthquake of June 16, 1964. Part 2. Estimation of earthquake moment, released energy, and stress-strain drop from the G-wave spectrum. *Bull. Earthq. Res. Inst., Tokyo Univ.* **44**, 73–88 (1966).
11. Wiens, D. A., Anandakrishnan, S., Nyblade, A. & Aleqabi, G. Remote detection and monitoring of glacial slip from Whillans Ice Stream using seismic Rayleigh waves recorded by the TAMSEIS array. *Eos* **87**, abstr.-S44A-04 (2006).
12. Joughin, I. *et al.* Continued deceleration of Whillans Ice Stream, West Antarctica. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024319 (2005).
13. Bindschadler, R. A., Stephenson, S. N., MacAyeal, D. R. & Shabtaie, S. Ice dynamics at the mouth of Ice Stream B, Antarctica. *J. Geophys. Res.* **92**, 8885–8894 (1987).
14. Alley, R. B. In search of ice-stream sticky spots. *J. Glaciol.* **39**, 447–454 (1993).
15. Peters, M. E., Blankenship, D. D. & Morse, D. L. Analysis techniques for coherent airborne radar sounding: Application to West Antarctic ice streams. *J. Geophys. Res.* **110**, doi:10.1029/2004JB003222 (2005).
16. Das, S. & Aki, K. Fault planes with barriers: A versatile earthquake model. *J. Geophys. Res.* **82**, 5648–5670 (1977).
17. Lay, T. & Kanamori, H. in *Earthquake Prediction: An International Review* (eds Simpson, D. W. & Richards, P. G.) 579–592 (American Geophysical Union, Washington DC, 1981).
18. Anandakrishnan, S., Voigt, D. E., Alley, R. B. & King, M. A. Ice stream D flow speed is strongly modulated by the tide beneath the Ross Ice Shelf. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016329 (2003).
19. Russell, D. R. Development of a time-domain, variable-period surface-wave magnitude measurement procedure for application at regional and teleseismic distances. *Bull. Seismol. Soc. Am.* **96**, 665–677 (2006).
20. Beroza, G. C. & Jordan, T. H. Searching for slow and silent earthquakes using free oscillations. *J. Geophys. Res.* **95**, 2485–2510 (1990).
21. Rhie, J. & Romanowicz, B. Excitation of the Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling. *Nature* **431**, 552–556 (2004).
22. Kanamori, H. & Given, J. W. Use of long-period surface waves for rapid determination of earthquake source parameters. *Phys. Earth Planet. Inter.* **27**, 8–31 (1981).
23. Suda, N., Kazunari, K. & Fukao, Y. Earth's background free oscillations. *Science* **279**, 2089–2091 (1998).
24. Horgan, H. J. & Anandakrishnan, S. Static grounding lines and dynamic ice streams: Evidence from the Siple Coast, West Antarctica. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL027091 (2006).
25. Padman, L., Fricker, H. A., Coleman, R., Howard, S. L. & Erofeeva, S. A new tidal model for the Antarctic ice shelves and seas. *Ann. Glaciol.* **34**, 247–254 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements GPS receivers for the TIDES project were supplied by the University NAVSTAR Consortium. Seismic data were obtained from the Data Management Center of the Incorporated Research Institutions for Seismology. This research was funded by the Office of Polar Programs, US National Science Foundation. M.A.K. was partially funded by a NERC (UK) research fellowship. We thank R. B. Alley, R. A. Bindschadler, H. Horgan, I. Joughin, L. Peters and D. E. Voigt for planning and carrying out the TIDES field deployment.

Author Contributions D.A.W. found the ice slip signals on the seismic records and carried out the seismic processing and modelling. D.A.W. also filtered the GPS time series and calculated the slip nucleation locations and times from the GPS records. S.A. and J.P.W. carried out the GPS fieldwork and calculated the displacement time series from the three-dimensional GPS data. M.A.K. processed the raw GPS data to obtain the three-dimensional displacement time series. All authors participated in the interpretation of the results and preparing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.A.W. (doug@wustl.edu).

METHODS

GPS data analysis. The GPS antennas were mounted on poles at a height 1.5–2 m above the surface. Data were logged using Trimble NetRS dual-frequency receivers. The GPS data had a sample rate of 0.1 Hz and were processed relative to a base station installed on nearby stationary ice, giving baselines of length 37–147 km. Base station coordinates were determined in the International Terrestrial Reference Frame 2000 using a Precise Point Positioning approach²⁶. Kinematic Precise Point Positioning solutions confirmed that the base station was moving to a negligible degree. When positioning the ice stream sites, data for each baseline were analysed in turn using the Track software²⁷, solving for three-dimensional WIS site coordinates and tropospheric zenith delay parameters at the full data rate. Carrier phase ambiguities were fixed to integers where possible. GPS satellite positions were adopted from the International GNSS Service final, precise orbit products²⁸. In relative positioning over these baseline lengths, ocean tide loading displacements and other unmodelled large-scale loading displacements are reduced to negligible levels. Data were processed in 28-h windows centred on 12:00 UT to reduce day boundary jumps and the coordinate time series were subsequently truncated to cover a single day. Site motion was loosely constrained at 0.05 m every 10 s, which reduces noise without over-smoothing the signal. The precision of the coordinates was assessed during WIS inter-slip time periods or at linearly moving sites and is ~ 0.01 m and ~ 0.03 m in the horizontal and vertical coordinates, respectively.

One-dimensional positions were calculated from the resulting GPS horizontal-coordinate time series using vector addition. These positions were filtered using a 600-s low-pass filter to reduce temporally correlated noise, and differentiated to produce velocities (Fig. 2). The nucleation points of the slip events were found by picking the first point of the filtered time series that rises above the background noise level and solving for the best-fit origin point and constant rupture velocity using a grid-search method and a least-squares criterion. The rupture velocity was found to decrease, and the misfit to increase, if more distant sites were included, suggesting that rupture velocity decreases with distance from the origin; therefore, the final locations were determined using only the five GPS receivers nearest the origin point.

Seismic data analysis. The geographical source regions radiating the second and third surface-wave packets were determined by first constraining the average velocities for Love-wave propagation to QSPA and Rayleigh-wave propagation to VNDA using the initial packet, for which the source location is known from GPS measurements (see above). Then we performed a grid search over the WIS to find the location best fitting the seismic surface-wave arrival times as well as the average rupture velocity of the source, starting from the observed nucleation point and using a least-squares criterion.

We modelled the seismic waveforms assuming that the source can be represented by a double couple at a very shallow depth (1 km). Previous studies of long-period surface-wave radiation from glacial sources² have used a single-force formalism^{29,30}. However, double-couple and horizontal single-force solutions become identical for very shallow sources³¹; thus, either formalism should be valid. In this case some approximation in the modelled signal amplitude may be introduced by the assumption of there being similar shear moduli on both sides of the dislocation, but this approximation should be satisfactory because we do not attempt to make a precise interpretation of the signal amplitude in terms of source properties. The source is assumed to occur in the till layer near the base of the ice stream, but there is very little difference if the source is placed in the ice itself, as the shear moduli of ice and water-saturated till are similar³².

We used a reflectivity algorithm³³ to compute the synthetics. Because the path from the WIS to VNDA traverses mostly the Ross Sea, for the VNDA synthetics we assumed a crust and mantle structure that was determined from surface waves in the region of the Ross Sea around Ross Island³⁴. For the QSPA synthetics we used a seismic structure determined from surface waves in East Antarctica³⁴ with an appropriate thickness of ice, as the largest portion of the path traverses East Antarctic lithosphere near the South Pole.

26. Zumberge, J. F., Heflin, M. B., Jefferson, D. C., Watkins, M. M. & Webb, F. H. Precise point positioning for the efficient and robust analysis of GPS data from large networks. *J. Geophys. Res.* **102**, 5005–5017 (1997).
27. Chen, G. *GPS Kinematics Positioning for the Airborne Laser Altimetry at Long Valley, California*. PhD thesis, Massachusetts Institute of Technology (1998).
28. Dow, J. M., Neilan, R. E. & Gendt, G. The international GPS service: celebrating the 10th anniversary and looking to the next decade. *Adv. Space Res.* **36**, 320–326 (2005).
29. Eissler, H. K. & Kanamori, H. A single-force model for the 1975 Kalapana, Hawaii, earthquake. *J. Geophys. Res.* **92**, 4827–4836 (1987).
30. Kawakatsu, H. Centroid single force inversion of seismic waves generated by landslides. *J. Geophys. Res.* **94**, 12363–12374 (1989).
31. Dahlen, F. A. Single-force representation of shallow landslide sources. *Bull. Seismol. Soc. Am.* **83**, 130–143 (1993).
32. Anandakrishnan, S. & Winberry, J. P. Antarctic subglacial sedimentary layer thickness from receiver function analysis. *Global Planet. Change* **42**, 167–176 (2004).
33. Kennett, B. L. N. *Seismic Wave Propagation in Stratified Media* (Cambridge Univ. Press, Cambridge, UK, 1983).
34. Lawrence, J. F. *et al.* Rayleigh wave phase velocity analysis of the Ross Sea, Transantarctic Mountains, and East Antarctica from a temporary seismic deployment. *J. Geophys. Res.* **111**, doi:10.1029/2005GL024523 (2006).

Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution

Rolando González-José¹, Ignacio Escapa², Walter A. Neves³, Rubén Cúneo² & Héctor M. Pucciarelli⁴

Evolutionary novelties in the skeleton are usually expressed as changes in the timing of growth of features intrinsically integrated at different hierarchical levels of development¹. As a consequence, most of the shape-traits observed across species do vary quantitatively rather than qualitatively², in a multivariate space³ and in a modularized way^{4,5}. Because most phylogenetic analyses normally use discrete, hypothetically independent characters⁶, previous attempts have disregarded the phylogenetic signals potentially enclosed in the shape of morphological structures. When analysing low taxonomic levels, where most variation is quantitative in nature, solving basic requirements like the choice of characters and the capacity of using continuous, integrated traits is of crucial importance in recovering wider phylogenetic information. This is particularly relevant when analysing extinct lineages, where available data are limited to fossilized structures. Here we show that when continuous, multivariate and modularized characters are treated as such, cladistic analysis successfully solves relationships among main *Homo* taxa. Our attempt is based on a combination of cladistics, evolutionary-development-derived selection of characters, and geometric morphometrics methods. In contrast with previous cladistic analyses of hominid phylogeny, our method accounts for the quantitative nature of the traits, and respects their morphological integration patterns. Because complex phenotypes are observable across different taxonomic groups and are potentially informative about phylogenetic relationships, future analyses should point strongly to the incorporation of these types of trait.

Cladistic analysis provides a solid framework to reconstruct phylogenetic relationships among taxa, because it identifies monophyletic groups by looking for shared derived characters. Theoretically, most cladistic methods need these characters to be discrete and independent, among other requirements. However, completion of cladistic analysis becomes problematic because, at lower taxonomic levels, most of observable variation is expressed as continuous changes of size and shape^{2,3} rather than in discrete identifiable structures. Although some traits can be reasonably treated as discrete, it is also true that modern morphometrics provide a rich source of quantitative characters, which raises the question of how to use them in inferring phylogenies. This problem has some important implications. First, apart from the fact that discretization methods are still the subject of intense debate^{2,7}, these procedures disregard the continuous nature of many complex morphological traits. Note that discretization can be either explicit, through a broad spectrum of gap-weighting methods applied on an admittedly continuous trait, or implicit through arbitrary definition of discrete character states upon a complex trait. In this case, the morphological features can be

described quantitatively, but are presented qualitatively (for example, low position of the infraorbital foramen). Second, discretization procedures force the multivariate nature of many complex phenotypes to be artificially treated as a collection of univariate measurements, disregarding the multivariate and geometric nature of form³. This is of crucial importance in cladistic practice, because homoplasy could be less likely in multivariate, complex phenotypes than in univariate traits⁸. Finally, at these low taxonomic levels, quantitative and developmental genetics can provide powerful additional tools to estimate degrees of character independence^{5,9}. In this context, functional and developmental integration leads to the co-inheritance of character complexes, often called modules, which are then constrained to evolve in a coordinated, rather than independent, fashion⁹. In human palaeontology, for instance, this principle is routinely violated as functionally and developmentally linked traits are subdivided for analytical purposes¹⁰. A logical approach to this fact in phylogenetic systematics is to treat integrated features as a single phylogenetic complex, and to treat the complex as if it were an independent character^{5,11}.

Here we present a cladistic analysis of the most complete fossil specimens pertaining to the hominid lineage, which explicitly takes into account the above implications. Our analysis considers the genus *Homo* as the ingroup, and the remaining specimens assigned to the genera *Gorilla*, *Pan*, *Australopithecus* and *Paranthropus* as outgroups (Table 1). The choice of characters is based on the most conservative approach in terms of modularity. The classical bulk of characters used previously¹² is reduced to just four modular characters that condense the main craniofacial shape changes observed in the analysed taxa. Geometric morphometrics¹³ methods are used to capture shape changes on these characters respecting both the geometric and the multivariate concept of shape. Finally, these geometric-morphometrics-derived, continuous, multivariate and modular characters are treated as such in a cladistic analysis.

Characters selected for analysis are flexure of the cranial base, facial retraction, neurocranial globularity, and shape and relative position of the masticatory apparatus. These characters reflect principal trends of variation on structures behaving as modules by varying somewhat independently^{11,14}. Even when further localized modules can be detected or hypothesized, our attempt here is to evaluate the phylogenetic signal contained in this restricted modularity hypothesis. Furthermore, these traits reflect the major evolutionary trends that acted to differentiate the hominid lineage. Even though multiple characters reflecting aspects of these traits have been cladistically analysed in previous studies, they have not been used in a modular way in a phylogenetic-systematic framework. Three-dimensional landmarks reflecting the shape of the modular characters were

¹Unidad de Investigación de Diversidad, Sistemática y Evolución, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Boulevard Brown 2825, U9120ACF Puerto Madryn, Argentina. ²Museo Paleontológico Egidio Feruglio, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Avenida Fontana 140, U9100GYO, Trelew, Argentina. ³Laboratório de Estudos Evolutivos Humanos, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, CP 11461, 05422-970 São Paulo, Brazil. ⁴Departamento Científico de Antropología del Museo de La Plata, Universidad Nacional de La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, 1900 La Plata, Argentina.

digitized on casts of the specimens listed in Table 1 (see Methods and Supplementary Information). Raw landmark coordinates were converted to shape coordinates by generalized Procrustes analysis. Generalized Procrustes analysis was performed for each module separately, and the aligned specimens were submitted to a principal component analysis¹³. Principal components can enable the pattern of variation between specimens described by many variables to be summarized by relatively few, when the data (for example, relative landmark locations) covary. The loading of each specimen on all the principal components necessary to achieve the 75% of variance explained was used as a continuous variable depicting the character state. Phylogenetic analysis of the resulting matrix was performed using the maximum parsimony algorithm⁶ for additive characters implemented in TNT¹⁵, and the maximum likelihood algorithm for quantitative traits developed by Felsenstein¹⁶ available in Phylip¹⁷ (see Methods).

One tree of maximum parsimony was obtained using a heuristic search, with 10,000 random addition sequences, saving 10 trees per replicate (Fig. 1a). In addition, one tree of maximum likelihood was also computed (Fig. 1b). The monophyletic status of the genus *Homo* is the most remarkable result in both analyses. Hypothetical ancestral character states for the *Homo* clade (node 8) are presented in Fig. 2 and Supplementary Information. With respect to the outgroups, this clade shows a more flexed cranial base, more retracted faces and an increase in the neurocranial globularity. This particular morphology is usually used to define the genus as well as to discuss the inclusion of some taxa in it^{18,19}.

Moreover, the internal relationships of the *Homo* specimens show a remarkable agreement with previous phylogenetic hypotheses²⁰. Maximum parsimony and maximum likelihood analyses only differ in the relative position of *H. sapiens* in relation to the complex *H. erectus*, *H. ergaster*, *H. rhodesiensis*. Whereas the maximum parsimony cladogram places *H. sapiens* as a sister of a clade formed by specimens assigned to *H. erectus*, *H. ergaster* and *H. rhodesiensis*, maximum likelihood accommodates *H. sapiens* in a derived position relative to it. What is coincident in both analyses is the association of two controversial specimens (D2700 and Broken Hill) to a clade also formed by *H. erectus* and *H. ergaster*, as previously suggested^{21–23}. *H. neanderthalensis* and *H. heidelbergensis* are represented in our ana-

lyses by several specimens, which form a single separate monophyletic group. Thus, our results are in agreement with previous assertions^{23,24} recognizing *H. heidelbergensis* and *H. neanderthalensis* as chronological variants inside a single biological lineage. The fact that *H. neanderthalensis sensu stricto* does not form a monophyletic clade with *H. sapiens* reinforces the idea that they are separate species. Conceptually, this is a key support for the method presented here, because this observation is also defended by studies based in evidence other than skull shape, such as analyses of ancient DNA²⁵ and growth patterns²⁶.

Finally, *H. rudolfensis* is the sister group of all the previous clades, and *H. habilis* is at the base of the monophyletic *Homo* clade (Fig. 1a). Based on our matrix, and as previously stated²⁷, there is no convincing reason to remove *H. habilis* and *H. rudolfensis* from the genus¹⁸, and their position goes counter to their inclusion in the *Australopithecus* genus¹⁹. In addition, our analysis supports the notion that *H. habilis* and *H. erectus* indeed represent different lineages even though they were recently reported as sympatric and contemporaneous forms²⁸.

Even when the focus of this work is on the relationships within *Homo*, the outgroup topology also shows some relevant points. For

Table 1 | Fossil (top) and recent (bottom) samples included in the analysis*

Specimen	Species assigned	Known age range (Myr)	Code
A.L. 444-2 (reconstruction)	<i>Australopithecus afarensis</i>	≥3.7–3.0	Aafa
Sts 5	<i>A. africanus</i>	ca. 3.0–2.5	Aafr
KNMER-406	<i>Paranthropus boisei</i>	≥2.3–1.4	Pboi-406
OH 5	<i>P. boisei</i>	≥2.3–1.4	Pboi-OH5
SK 48	<i>P. robustus</i>	ca. 1.5–2.0	Prob
WT 17000	<i>P. aethiopicus</i>	ca. 2.7–2.3	Paet
KNMER 1470	<i>Homo rudolfensis</i>	2.5–1.8	Hrud
KNMER 1813	<i>H. habilis</i>	2.1–1.5	Hhab
KNMER 3733	<i>H. ergaster</i>	2–1	Herg
Zhoukoudian†	<i>H. erectus</i>	1.8–0.03	Here
D2700	<i>H. erectus/H. ergaster</i>	1.8	Here-Herg
Steinheim	<i>H. heidelbergensis</i>	0.8–0.2	Hhei-S
Kabwe, Broken Hill 1	<i>H. rhodesiensis</i>	0.8–0.2	Hrho
Atapuerca 5	<i>H. heidelbergensis</i>	0.8–0.2	Hhei-A
Gibraltar 1, Forbes' Quarry	<i>H. neanderthalensis</i>	0.2–0.03	Hnea-G
La Chappelle-aux-Saints 1	<i>H. neanderthalensis</i>	0.2–0.03	Hnea-LC
La Ferrassie 1	<i>H. neanderthalensis</i>	0.2–0.03	Hnea-LF
CTL-004	<i>Gorilla gorilla</i>	9–0‡	Ggor
CTL-006	<i>Pan troglodytes</i>	8–0‡	Ptro
Patagonian, Rio Negro #797	<i>H. sapiens</i>	0.2–0	Hsap

* All specimens are stored at the Laboratorio de Estudos Evolutivos Humanos, University of São Paulo, Brazil, except Prob, Hgeo and Hhei-A, which are stored at the Unitat d'Antropologia, Faculty of Biology, University of Barcelona, and Hsap, which is stored at the Museo de La Plata, Argentina.

† First-generation casts were used for the fossil specimens. Sawyer and Tattersall's reconstruction.

‡ Estimated (molecular) time of disruption from the lineage of *H. sapiens*.

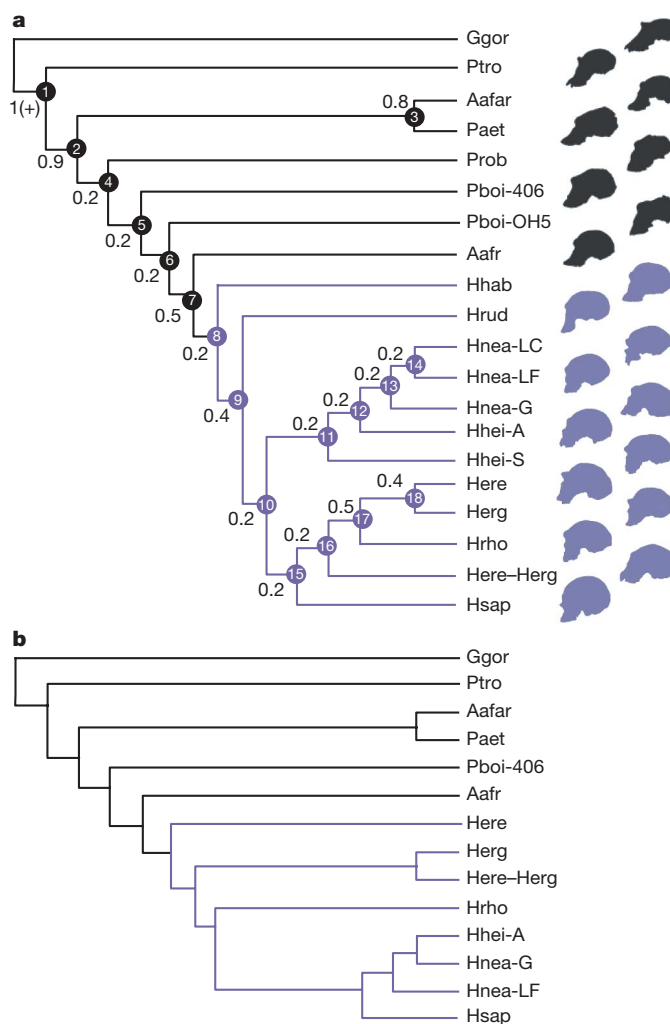


Figure 1 | Phylogenetic relationships among *Homo* species and other hominid taxa. a, Single tree obtained by equal weighted maximum-parsimony analysis based on morphological data of four cranial morphological modules. Bremer support values are displayed, as well as a numeric label for each node. The Bremer support values were determined by examination of the strict consensus of trees 0.01–0.12 steps longer than the shortest tree found for the data set (Supplementary Information). **b**, Single tree obtained by maximum likelihood. The ingroup (*Homo* specimens) and outgroups are displayed in purple and black respectively.

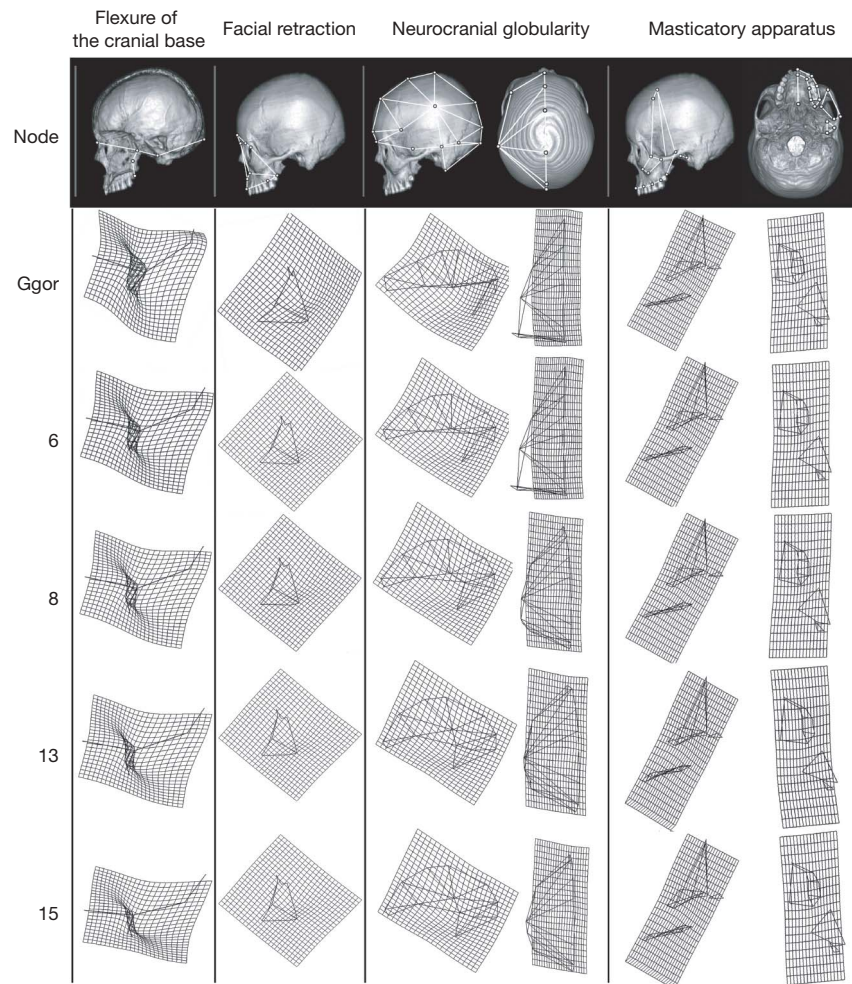


Figure 2 | Reconstruction of ancestral states corresponding to the root and the main nodes of the maximum parsimony cladogram. Ancestral states corresponding to the first principal component of each trait, estimated as values across the principal component of each character, and visualized as

deformation grids from the reference (the origin of the first principal component) towards the estimated principal component score of each node. Variation and ancestral states corresponding to further principal components can be explored using Supplementary Information.

instance, *Australopithecus afarensis* and *Paranthropus aethiopicus* are in a basal branch of the hominid clade, in a derived position respecting *Pan*, which is placed at the base of the tree, over the root (*Gorilla*). Neither maximum parsimony nor maximum likelihood cladograms support the genus status of *Paranthropus* or *Australopithecus*, because the fossils classically included in each of these genera fail to form monophyletic clades. Note that the paraphyletism of *Paranthropus* has been previously reported¹². Interestingly, our analyses place *Australopithecus africanus* as the sister group of the genus *Homo*.

In summary, our approach shows how classically disregarded information recovers significant evolutionary signals, using a new and promising methodology. Certainly, our tree recovered the monophyletic status of *Homo* as well as some of the most undisputed internal relationships (see a review in refs 20 and 22). However, our analysis is based on only four traits selected for previous knowledge about the relative independence of modules, and which treats the multivariant, geometric and continuous nature of skull shape as such.

The theoretical implications of the approach presented here are broad, but point in three main directions. First, reconstruction of phylogenetic relationships can be done by taking into account the modular development and evolution of complex phenotypes. In fact, this type of critical study on character independence should be the first phase of any cladistic analysis when no previous information is available⁵. Second, there is no reason to discard or force discretization of continuous multivariant traits. Conversely, our analysis shows that

it is possible to recover relevant phylogenetic signals in characters previously ignored or arbitrarily discretized. Note, however, that even though there are relatively few truly discrete traits, many of them exhibit qualitatively different states that are very distinct in different groups of taxa, and thus are reasonably discrete. In this context, future work should emphasize the combined use of quantitative and qualitative traits. Finally, when geometric morphometrics methods are used to depict shape changes across a multivariate space, reconstruction of ancestral states in combination with the visualization of shape changes across phyletic lineages should be considered as a straightforward and common-sense procedure. In summary, valuable phylogenetic information is recovered from data sets that consider independence on a developmentally and functionally basis, and which preserve the multivariant and continuous nature of complex phenotypes.

METHODS SUMMARY

Three-dimensional craniofacial landmark coordinates were digitized on 17 fossil hominid specimens and *Pan paniscus*, *Gorilla gorilla* and *Homo sapiens* skulls (Table 1). Landmarks were divided into four different subsets describing flexure of the cranial base, facial retraction, neurocranial globularity and the masticatory apparatus. Each subset was superimposed using generalized Procrustes analysis to remove the effects of translation, rotation and scaling, and then submitted to a principal component analysis. Projection of each specimen on the principal components was used as the shape descriptor and considered as a continuous trait in the cladistic analysis. Trees were obtained using *Gorilla gorilla* as the root,

and using maximum parsimony as well as maximum likelihood criteria. Ancestral states obtained from the maximum parsimony algorithm were used to visualize the shape changes corresponding to each node on the tree.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 October 2007; accepted 5 March 2008.

Published online 4 May 2008.

- Olson, E. C. & Miller, R. L. *Morphological Integration* (Univ. Chicago Press, Chicago, 1958).
- Rae, T. C. The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics* **14**, 221–228 (1998).
- MacLeod, N. & Forey, P. in *Morphology, Shape and Phylogeny* (eds MacLeod, N. & Forey, P.) 1–7 (Taylor & Francis, London, 2002).
- Pigliucci, M. & Preston, K. *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes* (Oxford Univ. Press, New York, 2004).
- Strait, D. S. Integration, phylogeny, and the hominid cranial base. *Am. J. Phys. Anthropol.* **114**, 273–297 (2001).
- Farris, J. S., Kluge, A. G. & Eckhardt, M. J. A numerical approach to phylogenetic systematics. *Syst. Zool.* **19**, 172–189 (1970).
- Humphries, C. J. in *Morphology, Shape and Phylogeny* (eds MacLeod, N. & Forey, P.) 8–26 (Taylor & Francis, London, 2002).
- Polly, P. D. On the simulation of the evolution of morphological shape under selection and drift. *Palaeontol. Electron.* **7**, 1–28 (2004).
- Cheverud, J. M. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *Am. Nat.* **145**, 63–89 (1995).
- Curnoe, D. Problems with the use of cladistic analysis in palaeoanthropology. *Homo* **53**, 225–234 (2003).
- Ackermann, R. R. & Cheverud, J. M. in *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes* (eds Pigliucci, M. & Preston, K.) 302–319 (Oxford Univ. Press, Oxford, 2004).
- Skelton, R. R. & McHenry, H. M. Evolutionary relationships among early hominids. *J. Hum. Evol.* **23**, 309–349 (1992).
- Zelditch, M. L., Swiderski, D. L., Sheets, H. D. & Fink, W. L. *Geometric Morphometrics for Biologists* (Elsevier, London, 2004).
- Lieberman, D. E., McBratney, B. M. & Krovitz, G. The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **99**, 1134–1139 (2002).
- Goloboff, P. A., Farris, J. S. & Nixon, K. TNT, a free program for phylogenetic analysis. *Cladistics*. (in the press).
- Felsenstein, J. *Inferring Phylogenies* (Sinauer Associates, Sunderland, Massachusetts, 2004).
- Felsenstein, J. PHYLIP (Phylogeny Inference Package) v.3.67 (Department of Genome Sciences, University of Washington, Seattle, 2007).
- Wood, B. A. & Collard, M. The human genus. *Science* **284**, 65–71 (1999).
- Collard, M. & Wood, B. A. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1575–1610 (Springer, Berlin and Heidelberg, 2007).
- Strait, D. S., Grine, F. E. & Fleagle, J. G. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1782–1806 (Springer, Berlin and Heidelberg, 2007).
- Vekua, A. et al. A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
- Rightmire, G. P. in *Handbook of Paleoanthropology* (eds Henke, W., Hardt, T. & Tattersall, I.) 1695–1715 (Springer, Berlin and Heidelberg, 2007).
- Stringer, C. B. in *Paleoclimate and Evolution with Emphasis on Human Origins* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 524–531 (Yale Univ. Press, New Haven, 1995).
- Arsuaga, J. L., Martínez, I., García, A. & Lorenzo, C. The Sima de los Huesos crania (Sierra de Atapuerca, Spain). A comparative study. *J. Hum. Evol.* **33**, 219–281 (1997).
- Lalueza-Fox, C. et al. Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula. *Mol. Biol. Evol.* **22**, 1077–1081 (2005).
- Ramirez Rozzi, F. V. & Bermudez De Castro, J. M. Surprisingly rapid growth in Neanderthals. *Nature* **428**, 936–939 (2004).
- Strait, D. S. & Grine, F. E. Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. *J. Hum. Evol.* **47**, 399–452 (2004).
- Spoor, F. et al. Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature* **448**, 688–691 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Goloboff, M. Hernández, C. Lalueza, N. Martínez-Abadías, D. Pol and F. Ramírez-Rozzi for reading and discussing previous versions of this paper. We also thank R. E. Ambrosetto, R. Nicoletti and B. Nicoletti for their assistance during this work. The program TNT is freely available, thanks to a subsidy from the Willi Hennig Society.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.G.J. (rolando@cenpat.edu.ar).

METHODS

Data acquisition and geometric morphometrics. Data were collected as three-dimensional coordinates of anatomical landmarks on the casts of specimens listed in Table 1. The landmark used in each character is provided in the Supplementary Information. All crania were measured by one observer (R.G.J.) using a Microscribe G2X digitizer. Because many of the fossil specimens were incomplete, some landmarks were reconstructed using anatomical information from the preserved surrounding areas. Missing bilateral landmarks on one side only were estimated by reflection. Scores were standardized to a mean of zero and a variance equal to the proportion of the variance explained by the corresponding principal component²⁹. Principal components have a biological meaning, as orthogonal dimensions of variance, even though that is not equivalent to the meaning of a character¹³. They are not likely to be characters in their own right because they are directions of variation that are constrained to be orthogonal (by definition), not directions of evolutionary change. However, data matrices submitted to cladistic analysis are not formed by characters represented by consecutive principal components of a single morphological structure, but by a collection of the first principal components representing the main trends of morphological change on four independent modules. Thus, overlapping of two or more specimens in a given-character principal component score does not mean that they are similar for a principal component on the remaining characters.

Cladistic analysis. The matrix is composed of 20 cranial specimens (Table 1) and 18 characters, which correspond to the firsts principal components of each module necessary to account for the 75% of explained variance. Maximum parsimony cladistic analysis used equal weighted maximum parsimony implemented in TNT^{15,30}, which allowed the use of continuous characters as such by optimizing them following the classical algorithm⁶ for additive characters³¹. When using continuous characters, a synapomorphy could be just a subtle change in one character. Thus, using continuous characters implies that there are an infinite number of character states. The most parsimonious tree was obtained using a heuristic search with 10,000 random addition sequences followed by tree-branching-regrafting. Characters were polarized, using *G. gorilla* to root the tree. Branch support was estimated using the Bremer method, using the suboptimal trees from 0.1 to 1 additional steps. Ancestral shapes are the optimizations of the maximum parsimony tree using the Wagner algorithm⁶. When using continuous traits, the usual output of this algorithm is a range of ancestral states, according to the maximum parsimony principle. Thus, to visualize the ancestral shape, we mapped back the transformations by taking the central value of the range and back-standardizing the standardized scores to recover the raw score values. These raw scores were used as the 'target' to obtain the wireframe depicting shape changes from the reference. Alternatively, because quantitative traits are expected to reverse direction in a Brownian-motion-like manner, a maximum likelihood tree was computed following Felsenstein^{16,17}.

29. Polly, P. D. Paleophylogeography: the tempo of geographic differentiation in marmots (*Marmota*). *J. Mammal.* **84**, 369–384 (2003).
30. Goloboff, P. A., Farris, J. S. & Nixon, K. TNT: Tree analysis using New Technology v.1.1 (Willi Hennig Society, New York, 2008).
31. Goloboff, P. A., Mattoni, C. I. & Quinteros, A. S. Continuous characters analyzed as such. *Cladistics* **22**, 589–601 (2007).

Understanding individual human mobility patterns

Marta C. González¹, César A. Hidalgo^{1,2} & Albert-László Barabási^{1,2,3}

Despite their importance for urban planning¹, traffic forecasting² and the spread of biological^{3–5} and mobile viruses⁶, our understanding of the basic laws governing human motion remains limited owing to the lack of tools to monitor the time-resolved location of individuals. Here we study the trajectory of 100,000 anonymized mobile phone users whose position is tracked for a six-month period. We find that, in contrast with the random trajectories predicted by the prevailing Lévy flight and random walk models⁷, human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations. After correcting for differences in travel distances and the inherent anisotropy of each trajectory, the individual travel patterns collapse into a single spatial probability distribution, indicating that, despite the diversity of their travel history, humans follow simple reproducible patterns. This inherent similarity in travel patterns could impact all phenomena driven by human mobility, from epidemic prevention to emergency response, urban planning and agent-based modelling.

Given the many unknown factors that influence a population's mobility patterns, ranging from means of transportation to job- and family-imposed restrictions and priorities, human trajectories are often approximated with various random walk or diffusion models^{7,8}. Indeed, early measurements on albatrosses⁹, followed by more recent data on monkeys and marine predators^{10,11}, suggested that animal trajectory is approximated by a Lévy flight^{12,13}—a random walk for which step size Δr follows a power-law distribution $P(\Delta r) \sim \Delta r^{-(1+\beta)}$, where the displacement exponent $\beta < 2$. Although the Lévy statistics for some animals require further study¹⁴, this finding has been generalized to humans⁷, documenting that the distribution of distances between consecutive sightings of nearly half-a-million bank notes is fat-tailed. Given that money is carried by individuals, bank note dispersal is a proxy for human movement, suggesting that human trajectories are best modelled as a continuous-time random walk with fat-tailed displacements and waiting-time distributions⁷. A particle following a Lévy flight has a significant probability to travel very long distances in a single step^{12,13}, which seems to be consistent with human travel patterns: most of the time we travel only over short distances, between home and work, whereas occasionally we take longer trips.

Each consecutive sighting of a bank note reflects the composite motion of two or more individuals who owned the bill between two reported sightings. Thus, it is not clear whether the observed distribution reflects the motion of individual users or some previously unknown convolution between population-based heterogeneities and individual human trajectories. Contrary to bank notes, mobile phones are carried by the same individual during his/her daily routine, offering the best proxy to capture individual human trajectories^{15–19}.

We used two data sets to explore the mobility pattern of individuals. The first (D_1) consisted of the mobility patterns recorded over

a six-month period for 100,000 individuals selected randomly from a sample of more than 6 million anonymized mobile phone users. Each time a user initiated or received a call or a text message, the location of the tower routing the communication was recorded, allowing us to reconstruct the user's time-resolved trajectory (Fig. 1a, b). The time between consecutive calls followed a 'bursty' pattern²⁰ (see Supplementary Fig. 1), indicating that although most consecutive calls are placed soon after a previous call, occasionally there are long periods without any call activity. To make sure that the obtained results were not affected by the irregular call pattern, we also studied a data set (D_2) that captured the location of 206 mobile phone users, recorded every two hours for an entire week. In both data sets, the spatial resolution was determined by the local density of the more than 10^4 mobile towers, registering movement only when the user moved between areas serviced by different towers. The average service area of each tower was approximately 3 km^2 , and over 30% of the towers covered an area of 1 km^2 or less.

To explore the statistical properties of the population's mobility patterns, we measured the distance between user's positions at consecutive calls, capturing 16,264,308 displacements for the D_1 and 10,407 displacements for the D_2 data set. We found that the distribution of displacements over all users is well approximated by a truncated power-law:

$$P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r/\kappa) \quad (1)$$

with exponent $\beta = 1.75 \pm 0.15$ (mean \pm standard deviation), $\Delta r_0 = 1.5 \text{ km}$ and cutoff values $\kappa|_{D_1} = 400 \text{ km}$ and $\kappa|_{D_2} = 80 \text{ km}$ (Fig. 1c, see the Supplementary Information for statistical validation). Note that the observed scaling exponent is not far from $\beta = 1.59$ observed in ref. 7 for bank note dispersal, suggesting that the two distributions may capture the same fundamental mechanism driving human mobility patterns.

Equation (1) suggests that human motion follows a truncated Lévy flight⁷. However, the observed shape of $P(\Delta r)$ could be explained by three distinct hypotheses: first, each individual follows a Lévy trajectory with jump size distribution given by equation (1) (hypothesis A); second, the observed distribution captures a population-based heterogeneity, corresponding to the inherent differences between individuals (hypothesis B); and third, a population-based heterogeneity coexists with individual Lévy trajectories (hypothesis C); hence, equation (1) represents a convolution of hypotheses A and B.

To distinguish between hypotheses A, B and C, we calculated the radius of gyration for each user (see Supplementary Information), interpreted as the characteristic distance travelled by user a when observed up to time t (Fig. 1b). Next, we determined the radius of gyration distribution $P(r_g)$ by calculating r_g for all users in samples D_1 and D_2 , finding that they also can be approximated with a truncated power-law:

$$P(r_g) = \left(r_g + r_g^0\right)^{-\beta_r} \exp(-r_g/\kappa_r) \quad (2)$$

¹Center for Complex Network Research and Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA. ²Center for Complex Network Research and Department of Physics and Computer Science, University of Notre Dame, Notre Dame, Indiana 46556, USA. ³Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA.

with $r_g^0 = 5.8$ km, $\beta_r = 1.65 \pm 0.15$ and $\kappa = 350$ km (Fig. 1d, see Supplementary Information for statistical validation). Lévy flights are characterized by a high degree of intrinsic heterogeneity, raising the possibility that equation (2) could emerge from an ensemble of identical agents, each following a Lévy trajectory. Therefore, we determined $P(r_g)$ for an ensemble of agents following a random walk (RW), Lévy flight (LF) or truncated Lévy flight (TLF) (Fig. 1d)^{8,12,13}. We found that an ensemble of Lévy agents display a significant degree of heterogeneity in r_g ; however, this was not sufficient to explain the truncated power-law distribution $P(r_g)$ exhibited by the mobile phone users. Taken together, Fig. 1c and d suggest that the difference in the range of typical mobility patterns of individuals (r_g) has a strong impact on the truncated Lévy behaviour seen in equation (1), ruling out hypothesis A.

If individual trajectories are described by an LF or TLF, then the radius of gyration should increase with time as $r_g(t) \sim t^{3/(2+\beta)}$ (ref. 21), whereas, for an RW, $r_g(t) \sim t^{1/2}$; that is, the longer we observe a user, the higher the chance that she/he will travel to areas not visited before. To check the validity of these predictions, we measured the time dependence of the radius of gyration for users whose gyration radius would be considered small ($r_g(T) \leq 3$ km), medium ($20 < r_g(T) \leq 30$ km) or large ($r_g(T) > 100$ km) at the end of our observation period ($T = 6$ months). The results indicate that

the time dependence of the average radius of gyration of mobile phone users is better approximated by a logarithmic increase, not only a manifestly slower dependence than the one predicted by a power law but also one that may appear similar to a saturation process (Fig. 2a and Supplementary Fig. 4).

In Fig. 2b, we chose users with similar asymptotic $r_g(T)$ after $T = 6$ months, and measured the jump size distribution $P(\Delta r | r_g)$ for each group. As the inset of Fig. 2b shows, users with small r_g travel mostly over small distances, whereas those with large r_g tend to display a combination of many small and a few larger jump sizes. Once we rescaled the distributions with r_g (Fig. 2b), we found that the data collapsed into a single curve, suggesting that a single jump size distribution characterizes all users, independent of their r_g . This indicates that $P(\Delta r | r_g) \sim r_g^{-\alpha} F(\Delta r / r_g)$, where $\alpha \approx 1.2 \pm 0.1$ and $F(x)$ is an r_g -independent function with asymptotic behaviour, that is, $F(x) \sim x^{-\alpha}$ for $x < 1$ and $F(x)$ rapidly decreases for $x \gg 1$. Therefore, the travel patterns of individual users may be approximated by a Lévy flight up to a distance characterized by r_g . Most important, however, is the fact that the individual trajectories are bounded beyond r_g ; thus, large displacements, which are the source of the distinct and anomalous nature of Lévy flights, are statistically absent. To understand the relationship between the different exponents, we note that the measured probability distributions are related

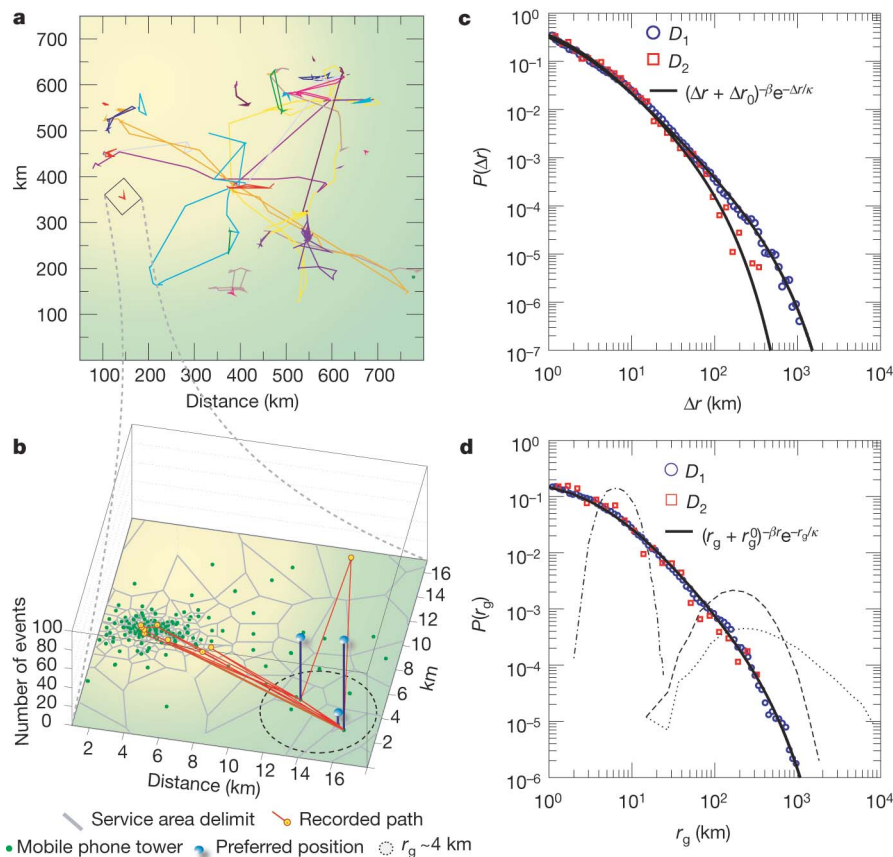


Figure 1 | Basic human mobility patterns. **a**, Week-long trajectory of 40 mobile phone users indicates that most individuals travel only over short distances, but a few regularly move over hundreds of kilometres. **b**, The detailed trajectory of a single user. The different phone towers are shown as green dots, and the Voronoi lattice in grey marks the approximate reception area of each tower. The data set studied by us records only the identity of the closest tower to a mobile user; thus, we can not identify the position of a user within a Voronoi cell. The trajectory of the user shown in **b** is constructed from 186 two-hourly reports, during which the user visited a total of 12 different locations (tower vicinities). Among these, the user is found on 96 and 67 occasions in the two most preferred locations; the frequency of visits

for each location is shown as a vertical bar. The circle represents the radius of gyration centred in the trajectory's centre of mass. **c**, Probability density function $P(\Delta r)$ of travel distances obtained for the two studied data sets D_1 and D_2 . The solid line indicates a truncated power law for which the parameters are provided in the text (see equation (1)). **d**, The distribution $P(r_g)$ of the radius of gyration measured for the users, where $r_g(T)$ was measured after $T = 6$ months of observation. The solid line represents a similar truncated power-law fit (see equation (2)). The dotted, dashed and dot-dashed curves show $P(r_g)$ obtained from the standard null models (RW, LF and TLF, respectively), where for the TLF we used the same step size distribution as the one measured for the mobile phone users.

by $P(\Delta r) = \int_0^\infty P(\Delta r | r_g) P(r_g) dr_g$, which suggests (see Supplementary Information) that up to the leading order we have $\beta = \beta_r + \alpha - 1$, consistent, within error bars, with the measured exponents. This indicates that the observed jump size distribution $P(\Delta r)$ is in fact the convolution between the statistics of individual trajectories $P(\Delta r | r_g)$ and the population heterogeneity $P(r_g)$, consistent with hypothesis C.

To uncover the mechanism stabilizing r_g , we measured the return probability for each individual $F_{pt}(t)$ (first passage time probability)^{21,22}, defined as the probability that a user returns to the position where he/she was first observed after t hours (Fig. 2c). For a two-dimensional random walk, $F_{pt}(t)$ should follow $\sim 1/(t \ln^2(t))$ (ref. 21). In contrast, we found that the return probability is characterized by several peaks at 24 h, 48 h and 72 h, capturing a strong tendency of humans to return to locations they visited before, describing the recurrence and temporal periodicity inherent to human mobility^{23,24}.

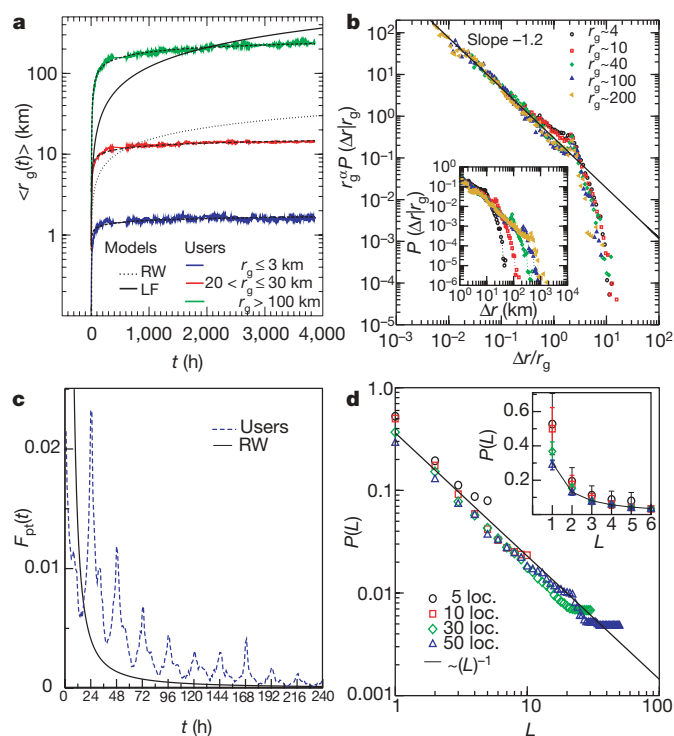


Figure 2 | The bounded nature of human trajectories. **a**, Radius of gyration $\langle r_g(t) \rangle$ versus time for mobile phone users separated into three groups according to their final $r_g(T)$, where $T = 6$ months. The black curves correspond to the analytical predictions for the random walk models, increasing with time as $\langle r_g(t) \rangle_{LF,TLF} \sim t^{3/2+\beta}$ (solid curve) and $\langle r_g(t) \rangle_{RW} \sim t^{0.5}$ (dotted curve). The dashed curves corresponding to a logarithmic fit of the form $A + B \ln(t)$, where A and B are time-independent coefficients that depend on r_g . **b**, Probability density function of individual travel distances $P(\Delta r | r_g)$ for users with $r_g = 4, 10, 40, 100$ and 200 km. As the inset shows, each group displays a quite different $P(\Delta r | r_g)$ distribution. After rescaling the distance and the distribution with r_g (main panel), the different curves collapse. The solid line (power law) is shown as a guide to the eye. **c**, Return probability distribution, $F_{pt}(t)$. The prominent peaks capture the tendency of humans to return regularly to the locations they visited before, in contrast with the smooth asymptotic behaviour $\sim 1/(t \ln^2(t))$ (solid line) predicted for random walks. **d**, A Zipf plot showing the frequency of visiting different locations (loc.). The symbols correspond to users that have been observed to visit $n_L = 5, 10, 30$ and 50 different locations. Denoting with (L) , the rank of the location listed in the order of the visit frequency, the data are well approximated by $R(L) \sim L^{-1}$. The inset is the same plot in linear scale, illustrating that 40% of the time individuals are found at their first two preferred locations; bars indicate the standard error.

To explore if individuals return to the same location over and over, we ranked each location on the basis of the number of times an individual was recorded in its vicinity, such that a location with $L = 3$ represents the third-most-visited location for the selected individual. We find that the probability of finding a user at a location with a given rank L is well approximated by $P(L) \sim 1/L$, independent of the number of locations visited by the user (Fig. 2d). Therefore, people devote most of their time to a few locations, although spending their remaining time in 5 to 50 places, visited with diminished regularity. Therefore, the observed logarithmic saturation of $r_g(t)$ is rooted in the high degree of regularity in the daily travel patterns of individuals, captured by the high return probabilities (Fig. 2b) to a few highly frequented locations (Fig. 2d).

An important quantity for modelling human mobility patterns is the probability density function $\Phi_a(x, y)$ to find an individual a in a given position (x, y) . As it is evident from Fig. 1b, individuals live and travel in different regions, yet each user can be assigned to a well defined area, defined by home and workplace, where she or he can be found most of the time. We can compare the trajectories of different users by diagonalizing each trajectory's inertia tensor, providing the probability of finding a user in a given position (see Fig. 3a) in the user's intrinsic reference frame (see Supplementary Information for the details). A striking feature of $\Phi(x, y)$ is its prominent spatial anisotropy in this intrinsic reference frame (note the different scales in Fig. 3a); we find that the larger an individual's r_g , the more pronounced is this anisotropy. To quantify this effect, we defined the anisotropy ratio $S \equiv \sigma_y / \sigma_x$, where σ_x and σ_y represent the standard deviation of the trajectory measured in the user's intrinsic reference frame (see Supplementary Information). We found that S decreases monotonically with r_g (Fig. 3c), being well approximated with $S \sim r_g^{-\eta}$ for $\eta \approx 0.12$. Given the small value of the scaling exponent, other functional forms may offer an equally good fit; thus, mechanistic models are required to identify if this represents a true scaling law or only a reasonable approximation to the data.

To compare the trajectories of different users, we removed the individual anisotropies, rescaling each user trajectory with its respective σ_x and σ_y . The rescaled $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ distribution (Fig. 3b) is similar for groups of users with considerably different r_g ; that is, after the anisotropy and the r_g dependence are removed all individuals seem to follow the same universal $\tilde{\Phi}(\tilde{x}, \tilde{y})$ probability distribution. This is particularly evident in Fig. 3d, where we show the cross section of $\tilde{\Phi}(x/\sigma_x, 0)$ for the three groups of users, finding that apart from the noise in the data the curves are indistinguishable.

Taken together, our results suggest that the Lévy statistics observed in bank note measurements capture a convolution of the population heterogeneity shown in equation (2) and the motion of individual users. Individuals display significant regularity, because they return to a few highly frequented locations, such as home or work. This regularity does not apply to the bank notes: a bill always follows the trajectory of its current owner; that is, dollar bills diffuse, but humans do not.

The fact that individual trajectories are characterized by the same r_g -independent two-dimensional probability distribution $\tilde{\Phi}(x/\sigma_x, y/\sigma_y)$ suggests that key statistical characteristics of individual trajectories are largely indistinguishable after rescaling. Therefore, our results establish the basic ingredients of realistic agent-based models, requiring us to place users in number proportional with the population density of a given region and assign each user an r_g taken from the observed $P(r_g)$ distribution. Using the predicted anisotropic rescaling, combined with the density function $\tilde{\Phi}(x, y)$, the shape of which is provided as Supplementary Table 1, we can obtain the likelihood of finding a user in any location. Given the known correlations between spatial proximity and social links, our results could help quantify the role of space in network development and evolution^{25–29} and improve our understanding of diffusion processes^{8,30}.

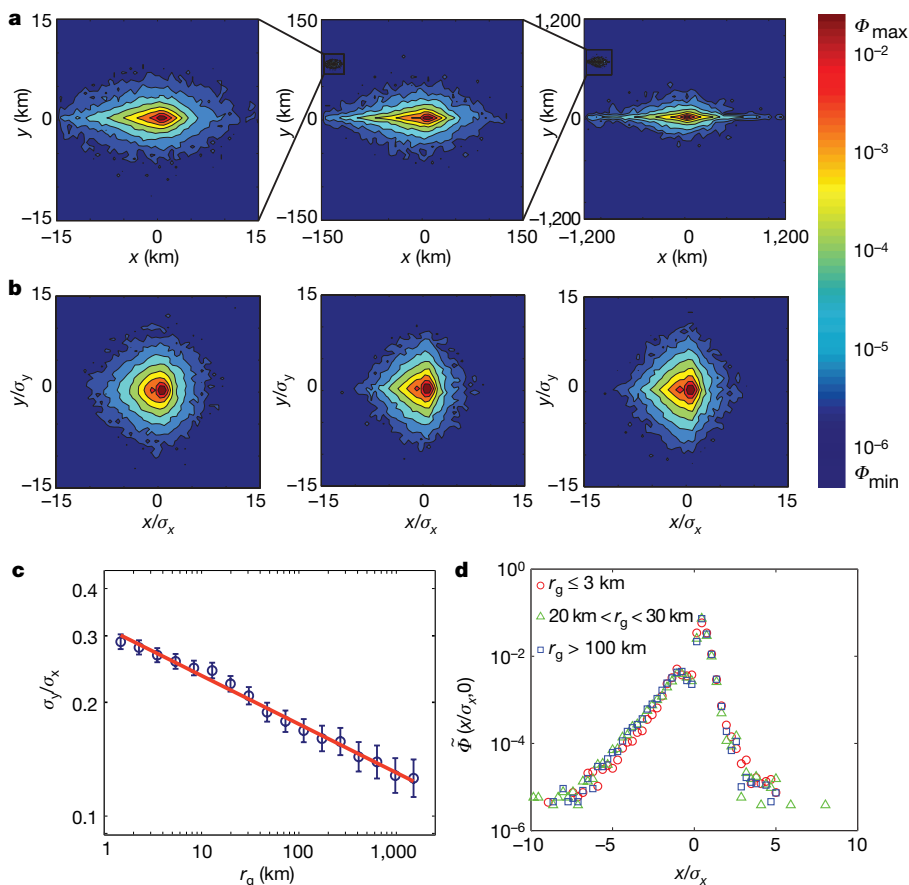


Figure 3 | The shape of human trajectories.

a, The probability density function $\Phi(x, y)$ of finding a mobile phone user in a location (x, y) in the user's intrinsic reference frame (see Supplementary Information for details). The three plots, from left to right, were generated for 10,000 users with: $r_g \leq 3$, $20 < r_g \leq 30$ and $r_g > 100$ km. The trajectories become more anisotropic as r_g increases. **b**, After scaling each position with σ_x and σ_y , the resulting $\Phi(x/\sigma_x, y/\sigma_y)$ has approximately the same shape for each group. **c**, The change in the shape of $\Phi(x, y)$ can be quantified calculating the isotropy ratio $S \equiv \sigma_y/\sigma_x$ as a function of r_g , which decreases as $S \sim r_g^{-0.12}$ (solid line). Error bars represent the standard error. **d**, $\tilde{\Phi}(x/\sigma_x, 0)$ representing the x -axis cross-section of the rescaled distribution $\Phi(x/\sigma_x, y/\sigma_y)$ shown in **b**.

Received 19 December 2007; accepted 27 March 2008.

- Horner, M. W. & O'Kelly, M. E. S Embedding economies of scale concepts for hub networks design. *J. Transp. Geogr.* **9**, 255–265 (2001).
- Kitamura, R., Chen, C., Pendyala, R. M. & Narayanan, R. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation* **27**, 25–51 (2000).
- Colizza, V., Barrat, A., Barthélémy, M., Valleron, A.-J. & Vespignani, A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Medicine* **4**, 95–110 (2007).
- Eubank, S. et al. Controlling epidemics in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15124–15129 (2004).
- Kleinberg, J. The wireless epidemic. *Nature* **449**, 287–288 (2007).
- Brockmann, D. D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
- Havlin, S. & Ben-Avraham, D. Diffusion in disordered media. *Adv. Phys.* **51**, 187–292 (2002).
- Viswanathan, G. M. et al. Lévy flight search patterns of wandering albatrosses. *Nature* **381**, 413–415 (1996).
- Ramos-Fernandez, G. et al. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behav. Ecol. Sociobiol.* **273**, 1743–1750 (2004).
- Sims, D. W. et al. Scaling laws of marine predator search behaviour. *Nature* **451**, 1098–1102 (2008).
- Klafter, J., Shlesinger, M. F. & Zumofen, G. Beyond brownian motion. *Phys. Today* **49**, 33–39 (1996).
- Mantegna, R. N. & Stanley, H. E. Stochastic process with ultraslow convergence to a gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* **73**, 2946–2949 (1994).
- Edwards, A. M. et al. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**, 1044–1049 (2007).
- Sohn, T. et al. in *Proc. 8th Int. Conf. UbiComp 2006* 212–224 (Springer, Berlin, 2006).
- Onnela, J.-P. et al. Structure and tie strengths in mobile communication networks. *Proc. Natl Acad. Sci. USA* **104**, 7332–7336 (2007).
- González, M. C. & Barabási, A.-L. Complex networks: from data to models. *Nature Physics* **3**, 224–225 (2007).
- Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
- Hidalgo, C. A. & Rodríguez-Sickert, C. The dynamics of a mobile phone network. *Physica A* **387**, 3017–3024 (2008).

- Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
- Redner, S. *A Guide to First-Passage Processes* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Condamine, S., Bénichou, O., Tejedor, V. & Klafter, J. First-passage times in complex scale-invariant media. *Nature* **450**, 77–80 (2007).
- Schlich, R. & Axhausen, K. W. Habitual travel behaviour: evidence from a six-week travel diary. *Transportation* **30**, 13–36 (2003).
- Eagle, N. & Pentland, A. Eigenbehaviours: identifying structure in routine. *Behav. Ecol. Sociobiol.* (in the press).
- Yook, S.-H., Jeong, H. & Barabási, A. L. Modeling the Internet's large-scale topology. *Proc. Natl Acad. Sci. USA* **99**, 13382–13386 (2002).
- Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology*. (Oxford Univ. Press, New York, 2007).
- Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, New York, 2003).
- Song, C. M., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **433**, 392–395 (2005).
- González, M. C., Lind, P. G. & Herrmann, H. J. A system of mobile agents to model social networks. *Phys. Rev. Lett.* **96**, 088702 (2006).
- Cecconi, F., Marsili, M., Banavar, J. R. & Maritan, A. Diffusion, peer pressure, and tailed distributions. *Phys. Rev. Lett.* **89**, 088102 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Brockmann, T. Geisel, J. Park, S. Redner, Z. Toroczkai, A. Vespignani and P. Wang for discussions and comments on the manuscript. This work was supported by the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems, the National Science Foundation within the DDDAS (CNS-0540348), ITR (DMR-0426737) and IIS-0513650 programs, and the US Office of Naval Research Award N00014-07-C. Data analysis was performed on the Notre Dame Biocomplexity Cluster supported in part by the NSF MRI grant number DBI-0420980. C.A.H. acknowledges support from the Kellogg Institute at Notre Dame.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.-L.B. (alb@neu.edu).

ADDENDUM

doi:10.1038/nature07850

Understanding individual human mobility patterns

Marta C. González, César A. Hidalgo & Albert-László Barabási

Nature 453, 779–782 (2008)

The human mobility study relied on an anonymized billing dataset that was previously recorded by a mobile provider as required by law and billing purposes, and not for the purposes of this project. The research was reviewed and approved by the Institutional Review Board (IRB) at the US Office of Naval Research, the main sponsor of the project, and has approval from the Northeastern University's IRB as well, the institution where the work was carried out. As part of the IRB review, the first author, who handled the data, and the PI participated in ethics training sessions at the outset of the study.

Dynamics of fat cell turnover in humans

Kirsty L. Spalding¹, Erik Arner¹, Pål O. Westermark², Samuel Bernard³, Bruce A. Buchholz⁴, Olaf Bergmann¹, Lennart Blomqvist⁵, Johan Hoffstedt⁵, Erik Näslund⁶, Tom Britton⁷, Hernan Concha⁵, Moustapha Hassan⁵, Mikael Rydén⁵, Jonas Frisén¹ & Peter Arner⁵

Obesity is increasing in an epidemic manner in most countries and constitutes a public health problem by enhancing the risk for cardiovascular disease and metabolic disorders such as type 2 diabetes^{1,2}. Owing to the increase in obesity, life expectancy may start to decrease in developed countries for the first time in recent history³. The factors determining fat mass in adult humans are not fully understood, but increased lipid storage in already developed fat cells (adipocytes) is thought to be most important^{4,5}. Here we show that adipocyte number is a major determinant for the fat mass in adults. However, the number of fat cells stays constant in adulthood in lean and obese individuals, even after marked weight loss, indicating that the number of adipocytes is set during childhood and adolescence. To establish the dynamics within the stable population of adipocytes in adults, we have measured adipocyte turnover by analysing the integration of ¹⁴C derived from nuclear bomb tests in genomic DNA⁶. Approximately 10% of fat cells are renewed annually at all adult ages and levels of body mass index. Neither adipocyte death nor generation rate is altered in early onset obesity, suggesting a tight regulation of fat cell number in this condition during adulthood. The high turnover of adipocytes establishes a new therapeutic target for pharmacological intervention in obesity.

The fat mass can expand by increasing the average fat cell volume and/or the number of adipocytes. Increased fat storage in fully differentiated adipocytes, resulting in enlarged fat cells, is well documented and is thought to be the most important mechanism whereby fat depots increase in adults^{4,5}. To analyse the contribution of the fat cell volume in adipocytes to the size of the fat mass, we first analysed the relationship between fat cell volume and total body fat mass (directly measured with bioimpedance or estimated from body mass index (BMI), sex and age in a large cohort of adults). As expected, there was a positive correlation between the measures of fat mass and fat cell volume both in subcutaneous fat (Fig. 1a–c), which represents about 80% of all fat, and in visceral fat (Fig. 1d), which has a strong link to metabolic complications of obesity. However, the relationship between fat cell volume and fat mass markedly differed from a linear relationship (likelihood ratio test $P < 0.001$, and Akaike information criterion, described in Supplementary Information 1) in both subcutaneous and visceral adipose regions and both sexes, indicating that fat mass is determined by both adipocyte number and size. In the nonlinear case, both fat cell number and fat cell size determine fat mass. If the relationship had been linear, fat cell volume would be the only important determinant of fat mass.

The generation of adipocytes is a major factor behind the growth of adipose tissue during childhood⁷, but it is unknown whether the number of adipocytes changes during adulthood. We assessed the

total adipocyte number in 687 adult individuals and combined this data with previously reported results for children and adolescents⁸. Although the total adipocyte number increased in childhood and adolescence, this number levelled off and remained constant in adulthood in both lean and obese individuals (adults over 20 yr, grouped in 5-yr bins; ANOVA, lean $P = 0.68$, obese $P = 0.21$; Fig. 2a and Supplementary Information 3). Thus, the difference in adipocyte number between lean and obese individuals is established during childhood^{7,8} and the total number of adipocytes for each weight category stays constant during adulthood (Fig. 2b). The small variation in adipocyte number for each BMI category demonstrates that this is a stable cell population during adulthood.

To analyse whether alterations in adipocyte number may contribute to changed fat mass under extreme conditions, we next asked whether fat cell number is reduced during major weight loss (mean body weight loss, $18 \pm 11\%$, mean \pm s.d.) by radical reduction in calorie intake by bariatric surgery (reduction of the stomach with the purpose of facilitating weight loss). The surgical treatment resulted in a significant decrease in BMI and fat cell volume; however, this failed to reduce adipocyte cell number two years post surgery (Fig. 2b, c and Supplementary Information 4), in line with previous studies using different methodology^{9–12}. Similar results were found in a complementary longitudinal study¹³. Ref. 13 found that significant weight gain (15–25%) over several months in non-obese adult men resulted in a significant increase in body fat, which was accompanied by an increase in adipocyte volume, but no change in adipocyte number. Similar to our findings, subsequent weight loss back to baseline resulted in a decrease of adipocyte volume, but, again, no change in adipocyte number. Although we cannot rule out that a more prolonged period of weight gain in adulthood could result in an increase in adipocyte number, these results and ours indicate that fat cell number is largely set by early adulthood and that changes in fat mass in adulthood can mainly be attributed to changes in fat cell volume. This may indicate that the number of adipocytes is set by early adulthood with no subsequent cell turnover. Alternatively, the generation of adipocytes may be balanced by adipocyte death, with the total number being tightly regulated and constant.

We next set out to establish whether adipocytes are replaced during adulthood, and, if so, at what rate. Adipocytes can be generated from adult human mesenchymal stem cells and pre-adipocytes *in vitro*¹⁴ and may undergo apoptosis or necrosis^{15–17}, but it is unclear whether adipocytes are generated *in vivo*¹⁴. Cell turnover has been difficult to study in humans. Methods used in experimental animals, such as the incorporation of labelled nucleotides, cannot readily be adapted for use in humans owing to potential toxicity. The detection of cells expressing molecular markers of proliferation can give

¹Department of Cell and Molecular Biology, Karolinska Institute, SE-171 77 Stockholm, Sweden. ²Institute for Theoretical Biology (ITB), Humboldt University Berlin and Charité, Invalidenstrasse 43, 10115 Berlin, Germany. ³Institute of Applied and Computational Mathematics, Foundation of Research and Technology, 71110 Heraklion Crete, Greece. ⁴Centre for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, 7000 East Avenue, L-397, Livermore, California 94551, USA. ⁵Department of Medicine, Karolinska University Hospital, SE-141 86 Stockholm, Sweden. ⁶Division of Surgery, Department of Clinical Science, Danderyds Hospital, Karolinska Institutet, SE-182 88 Stockholm, Sweden. ⁷Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden.

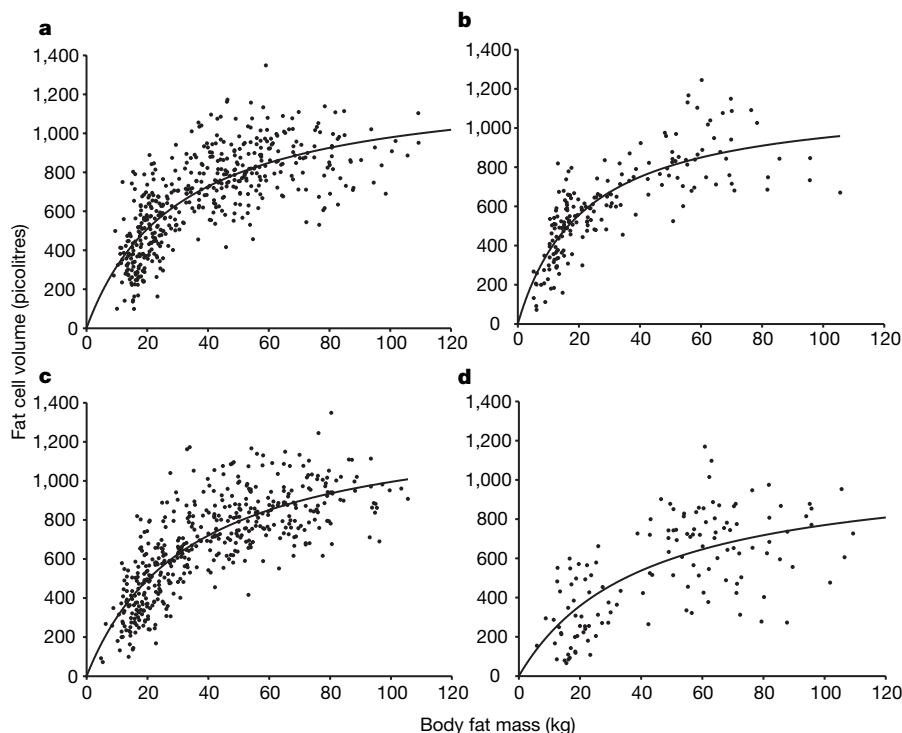


Figure 1 | Fat mass is determined by both adipocyte number and size.

a–d, The relationship between fat mass and fat cell volume was curvilinear across the range of body fat mass in female (**a**) and male (**b**) subcutaneous fat (n female = 480; n male = 190), in combined female and male subcutaneous fat (**c**; n female = 357, n male = 117), and in male and female visceral fat (**d**; n female = 84, n male = 51). This demonstrates that both adipocyte number

and adipocyte size are determinates of body fat mass. In **a**, **b** and **d**, body fat mass was estimated from BMI using a previously described formula (conversion formula is described in Supplementary Information 1 and 4); in **c**, fat mass was determined using bioimpedance. Fat cell volume is given in picolitres, where 10^{-12} litres = 10^{-9} cm³.

insights about mitotic activity, but fail to provide information regarding the fate of the progeny of the dividing cells. This is a limitation when studying postmitotic cell types, which do not divide or express mitotic markers themselves (for example, neurons or adipocytes) but may be replenished from proliferating stem or progenitor cells, such as preadipocytes.

In this study, we used a recently developed method that is based on the incorporation of ¹⁴C from nuclear bomb tests into genomic DNA and allows the analysis of cell turnover in humans^{6,18}. Levels of ¹⁴C in the atmosphere were relatively stable until the Cold War, when above-ground nuclear bomb tests (1955–1963) caused a notable increase^{19,20} (Fig. 3a, b). Even though the detonations were conducted at a limited number of locations, increased ¹⁴C levels in the atmosphere rapidly equalized around the globe. Since the Test-Ban Treaty in 1963, the ¹⁴C levels have dropped exponentially, not because of radioactive decay (half-life 5,730 yr), but by diffusion from the atmosphere²¹. Atmospheric ¹⁴C reacts with oxygen to form CO₂, which is

incorporated into plants by photosynthesis. By eating plants, and animals that live off plants, the ¹⁴C concentration in the human body closely parallels that in the atmosphere at any given point in time^{22–24}. Because DNA is stable after a cell has gone through its last cell division, the ¹⁴C level in DNA serves as a date mark for when a cell was born; this can be used to retrospectively birth-date cells in humans^{6,18}.

To address whether adipocytes are generated from newborn cells in adulthood, we isolated fat cells ($\geq 98\%$ purity, Supplementary Information 4) from adipose tissue collected during liposuction or abdominal wall reconstruction from 35 adult lean or obese individuals. The pure isolation of adipocytes is important because non-adipose cells are present in adipose tissue and these cell types may have a different turnover rate (see Supplementary Information 4 for a full discussion). Genomic DNA was extracted from the purified adipocytes, and ¹⁴C levels were measured by accelerator mass spectrometry and related to atmospheric ¹⁴C data (Fig. 3c, d and Supplementary Information 4). We first analysed individuals born

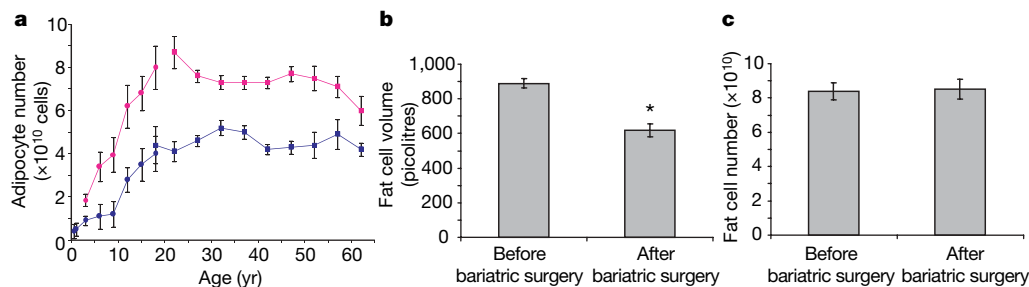


Figure 2 | Adipocyte number remains stable in adulthood, although significant weight loss can result in a decrease in adipocyte volume. Total adipocyte number from 595 (n lean = 253; n obese = 342) adult individuals (squares) was combined with previous results for children and adolescents⁸ (circles; n lean = 178; n obese = 120). **a**, The adipocyte number increases in

childhood and adolescence, with the number levelling off and remaining constant in adulthood in both lean (blue) and obese (pink) individuals. **b**, **c**, Major weight loss by bariatric surgery results in a significant decrease in cell volume (**b**), however fails to reduce adipocyte cell number (**c**), 1–2 yr post surgery ($n = 20$). All error bars represent s.e.m.; asterisk, $P < 0.0001$.

well before the period of nuclear bomb tests. This provides a high sensitivity to detect the generation of cells born after the time of onset of the nuclear bomb tests (1955), because ^{14}C levels above those present before the Cold War can be detected even if only a small (1%) proportion of cells in a population are renewed⁶. In all analysed individuals born before 1955 ($n = 10$), the ^{14}C levels were substantially higher than the atmospheric levels before the nuclear bomb tests, indicating that generation of adipocytes had taken place after 1955 (Fig. 3c, see Supplementary Information 2 for all ^{14}C measurements and associated data). The individuals were 0–22 years old at the onset of nuclear bomb tests, establishing that adipocytes are generated during adolescence and in early adulthood. New adipocytes may also be formed by differentiation of existing post-mitotic pre-adipocytes; hence, DNA integration of ^{14}C provides a lower bound to the generation of adipocytes.

Analysis of individuals born before the onset of the nuclear bomb tests provides a high sensitivity to detect cell turnover, but alone does not allow the establishment of the turnover rate because a certain ^{14}C level can correspond to the rising or the falling part of the atmospheric ^{14}C curve. However, the integration of data from individuals born before and after the period of nuclear bomb tests allows determination of cell turnover as well as the relative contribution of cell death and cell renewal to this process (see Supplementary Information 3). We therefore also analysed ^{14}C levels in adipocyte genomic DNA from individuals born after the period of nuclear bomb tests ($n = 25$). In all of these individuals, the ^{14}C levels corresponded to surprisingly contemporary time points (Fig. 3d and Supplementary Information 2), providing a first indication that there is continuous and substantial turnover of adipocytes in adult humans.

We next calculated the dynamics of fat cell turnover using a simple birth and death model (detailed in Supplementary Information 3). The model's assumptions allow the calculation of kinetic rates for individual subjects. The death rate of adipocytes is approximately $8.4 \pm 6.2\%$ per yr (median \pm average deviation) in the total fat pool of the body. The distribution of death rates is skewed towards lower values and is not the normal gaussian (Jarque–Bera test for normality, $P < 0.05$); therefore, the median \pm average deviation is more informative than the mean \pm s.d.²⁵. To test the reliability of the death-rate estimates, we used three different scenarios concerning the generation of adipocytes early in life, and confirmed that different estimates of the death rates do not differ from the median (sign test, $P > 0.3$; see Supplementary Information 3 for description of the scenarios). We divided the data set into lean ($\text{BMI} < 25 \text{ kg per m}^2$) and obese ($\text{BMI} \geq 30 \text{ kg per m}^2$, all of which had early onset obesity, see Supplementary Information 4) for analyses of the influence of obesity on adipocyte death rate. No significant difference in adipocyte death rate was seen across the different BMIs, with obese individuals having a median adipocyte death rate of $9.5 \pm 5.1\%$ (median \pm average deviation) per yr, versus $8.2 \pm 5.3\%$ (median \pm average deviation) per yr for lean individuals ($P = 0.6$ using the Kruskal–Wallis test, which tests for equality of medians; Fig. 4a). We found no trend for an increase in average cell number in subjects aged 20–70 yr using data presented in Fig. 2b ($n = 650$ and $P = 0.19$ by linear regression analysis), arguing that the adipocyte death rate per yr must be matched with a similar birth rate. This translates into an adipocyte turnover rate similar for all weight categories. We calculate a median turnover rate of $8.4 \pm 6.2\%$ (median \pm average deviation) per yr, with half of the adipocytes replaced every 8.3 yr.

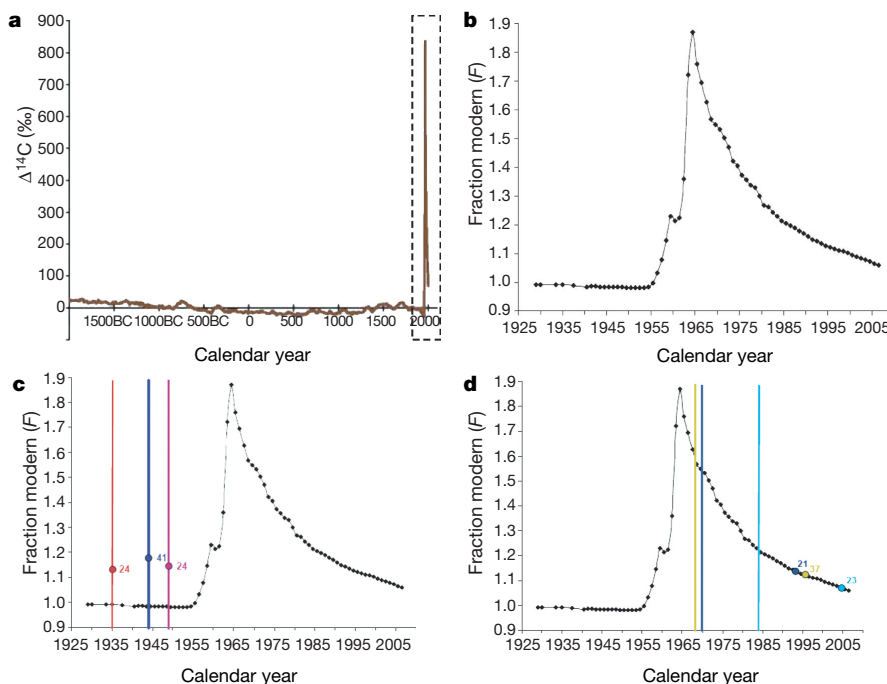


Figure 3 | Turnover of adipocytes in adulthood. **a, b**, The levels of ^{14}C in the atmosphere have been relatively stable over long time periods, with the exception of a large addition of ^{14}C in 1955–1963 as a result of nuclear bomb tests²¹. The boxed region in **a** is shown in more detail in **b**. ^{14}C levels from modern samples are by convention given in relation to a universal standard and corrected for radioactive decay, giving the $\Delta^{14}\text{C}$ value³⁰. **c, d**, Adipocyte age in adult human subjects born before (**c**) and after (**d**) nuclear bomb tests were analysed by determining the ^{14}C concentration in adipocyte genomic DNA using accelerator mass spectrometry. The measured ^{14}C value is related to the recorded atmospheric levels to establish at what time point

they corresponded. The year is plotted on the x axis, giving the birth date of the cell population. Three representative individuals born at different times before the onset of the bomb tests reveals the generation of adipocytes after birth (**c**). Analysis of the oldest individuals established that adipocytes are born in adolescence and in adulthood (**c**). ^{14}C levels analysed in people born after the period of nuclear bomb tests showed continuous and substantial turnover of adipocytes in adult humans (**d**). The time of birth of the person is indicated by a vertical line in each graph and the BMI is shown numerically (**c, d**). Error bars for the accelerator mass spectrometry readings are too small to be visualized in this graph. Each dot represents one individual.

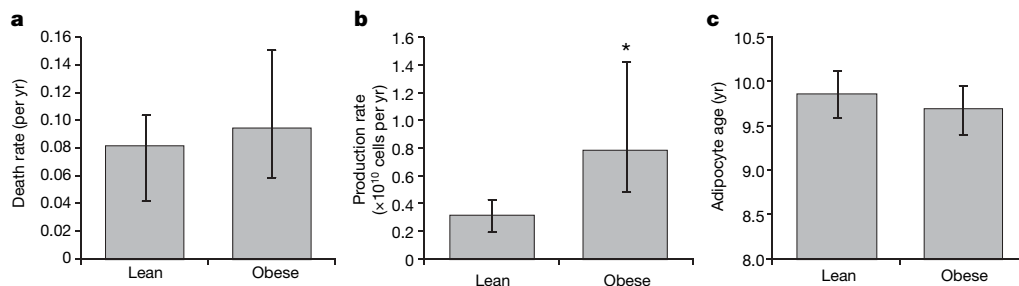


Figure 4 | Effect of obesity on adipocyte generation and death. **a**, No significant difference in adipocyte death rate per year was seen across the different BMIs. **b**, Obese individuals had a significantly greater number of adipocytes added per year than lean individuals. **c**, No significant difference in

the average age of adipocytes in lean versus obese individuals was found. In **a** and **b**, values are the medians and error bars indicate the location of the first and third quartiles; in **c**, data are shown as mean \pm s.e.m. Asterisk, $P < 0.01$ for lean ($n = 13$) versus obese ($n = 14$) individuals, Kruskal–Wallis test.

Using the death-rate estimates and the fat cell numbers calculated for individual subjects, absolute fat cell production was calculated. Obese individuals were found to have a significantly greater number of adipocytes added per year than lean individuals: $(0.8 \pm 0.5) \times 10^{10}$ cells per yr versus $(0.3 \pm 0.2) \times 10^{10}$ cells per yr (median \pm average deviation; $P < 0.01$, ANOVA; Fig. 4b). Loss of fat cells is therefore compensated by the production of new fat cells, which is twice as high in obese subjects compared with lean subjects. The fact that the total number of new adipocytes added each year is greater in obese compared with lean individuals, yet the proportion of newborn adipocytes added each year (the turnover) is the same for both groups, argues that the difference in cell number between the lean and obese adults occurs before adulthood. In support of this, we found no significant difference in the average age of adipocytes in lean 9.9 ± 3.5 yr (mean \pm s.d.) versus obese 9.7 ± 4.0 yr (mean \pm s.d.) individuals (Fig. 4c). No significant correlation between the age of subjects and cell death or between the age of patients and adipocyte generation was found (Supplementary Information 3), suggesting a constant turnover rate throughout adult life.

If the number of adipocytes is set to a higher level in obese people before adulthood, this could be because cell-number expansion begins earlier (age of onset), because expansion is faster (growth relative to the initial cell number (IC) at age of onset), or because expansion ends later (age at 90% of adult cell number). We used combined adipocyte number data (Fig. 2a) to see whether one or more of these factors determine adipocyte number. Using our birth and death model, we determined that age at onset of adipocyte number expansion is significantly earlier in obese (2.1 ± 0.9 yr) than in lean (5.7 ± 0.8 yr) subjects; the relative increase in adipocyte number is higher in obese (2.4 ± 0.6 IC yr $^{-1}$) than in lean (1.3 ± 0.3 IC yr $^{-1}$) subjects, but end of expansion of adipocyte number is earlier in obese (16.5 ± 1.3 yr) than in lean (18.5 ± 0.7 yr) subjects (all values are predicted values \pm 95% confidence interval, Supplementary Information 3). Therefore, adult cell number is set earlier in obese subjects and is not caused by a prolonged expansion period in adulthood.

We find that the number of adipocytes for lean and obese individuals is set during childhood and adolescence, and that adipocyte numbers for these categories are subject to little variation during adulthood. Even after significant weight loss in adulthood and reduced adipocyte volume, the adipocyte number remains the same. Although we show that the adipocyte number is static in adults, we also demonstrate that there is remarkable turnover within this population, indicating that adipocyte number is tightly controlled and not influenced by the energy balance. Studies of previously obese individuals after weight loss show that their adipose tissue hypercellularity is associated with leptin deficiency, which is likely to increase appetite and to lower energy expenditure²⁶. These factors promote lipid accumulation in fat cells and weight gain towards the status before weight loss. Thus, a tight regulation of adipocyte number, together with mechanisms maintaining their energy balance, may contribute to why obese individuals have difficulties maintaining weight loss.

It should be stressed that our conclusions on the rates of adipocyte turnover (^{14}C data) were obtained from studies on subjects with early onset of obesity. We cannot rule out that those who gradually gain significant weight over years in adulthood may initially increase their adipocyte size until a threshold is reached and thereafter recruit new fat cells from committed precursor cells or mesenchymal stem cells. Most obese adults have been obese since childhood, with less than 10% of children with normal weight going on to develop adult obesity²⁷. By contrast, over three-quarters of obese children go on to become obese adults²⁷. Thus, understanding the dynamics of adipocyte turnover in adults who have been obese since childhood is of great importance, especially given the current trend for an increase in childhood obesity.

The size of organs can be regulated by different mechanisms, and the number of cells in some tissues is controlled by a systemic feedback mechanism²⁸. This is best understood for skeletal muscle, in which growth and differentiation factor 8 (GDF8), also known as myostatin, is secreted from myocytes and negatively regulates the generation of new muscle cells and thereby sets the number of cells²⁹. Loss-of-function mutations in *GDF8* result in a large increase in the number (and size) of myocytes in animals and humans²⁹. The steady production of adipocytes in adults results in a stable size of the constantly turning over adipocyte population. Feedback mechanisms that control adipocyte turnover will be important to identify at a molecular level because this may offer a novel target for pharmacological therapy when obesity is established and for other types of intervention during childhood and adolescence when the final number of fat cells in the body is being set.

METHODS SUMMARY

Subjects. The relationship between subcutaneous or visceral fat cell volume, BMI and fat mass was studied in two separate cohorts, and fat cell turnover was studied in a third cohort, all of which are described in Supplementary Information 4.

Isolated fat cells. Fat cells were isolated from the adipose tissue as described in Supplementary Information 4. Details on how to measure weight, volume and the number of fat cells as well as determination of the purity of the adipocytes are given in Supplementary Information 4.

^{14}C analysis. Genomic DNA was prepared from isolated fat cells, and was purified and subjected to accelerator mass spectrometry analyses, as described in Supplementary Information 4 and tabled in Supplementary Information 2.

Data analysis. The calculations of relationship between fat cell volume and BMI or fat mass are described in detail in Supplementary Information 1. The calculations of fat cell death and generation are described in detail in Supplementary Information 3.

Received 30 November 2007; accepted 7 March 2008.

Published online 4 May 2008.

1. Van Gaal, L. F., Mertens, I. L. & De Block, C. E. Mechanisms linking obesity with cardiovascular disease. *Nature* **444**, 875–880 (2006).
2. Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840–846 (2006).
3. Olshansky, S. J. *et al.* A potential decline in life expectancy in the United States in the 21st century. *N. Engl. J. Med.* **352**, 1138–1145 (2005).

4. Bjorntorp, P. Effects of age, sex, and clinical conditions on adipose tissue cellularity in man. *Metabolism* **23**, 1091–1102 (1974).
5. Hirsch, J. & Batchelor, B. Adipose tissue cellularity in human obesity. *Clin. Endocrinol. Metab.* **5**, 299–311 (1976).
6. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisen, J. Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005).
7. Prins, J. B. & O'Rahilly, S. Regulation of adipose cell number in man. *Clin. Sci. (Lond.)* **92**, 3–11 (1997).
8. Knittle, J. L., Timmers, K., Ginsberg-Fellner, F., Brown, R. E. & Katz, D. P. The growth of adipose tissue in children and adolescents. Cross-sectional and longitudinal studies of adipose cell number and size. *J. Clin. Invest.* **63**, 239–246 (1979).
9. Miller, M. *et al.* Demonstration of *de novo* production of adipocytes in adult rats by biochemical and radioautographic techniques. *J. Lipid Res.* **25**, 336–347 (1984).
10. Kral, J. *et al.* Body composition and adipose tissue cellularity before and after jejuno-ileostomy in severely obese subjects. *Eur. J. Clin. Inv.* **7**, 413–419 (1977).
11. Bjorntorp, P. *et al.* Effect of an energy reduced dietary regimen in relation to adipose tissue cellularity in obese women. *Am. J. Clin. Nutr.* **28**, 445–452 (1975).
12. Häger, A. *et al.* Adipose tissue cellularity in obese school girls before and after dietary intervention. *Am. J. Clin. Nutr.* **31**, 68–75 (1978).
13. Sims, E. A. *et al.* Experimental obesity in man. *Trans. Assoc. Am. Physicians* **81**, 153–170 (1968).
14. Rodriguez, A. M., Elabd, C., Amri, E. Z., Ailhaud, G. & Dani, C. The human adipose tissue is a source of multipotent stem cells. *Biochimie* **87**, 125–128 (2005).
15. Petruschke, T. & Hauner, H. Tumor necrosis factor- α prevents the differentiation of human adipocyte precursor cells and causes delipidation of newly developed fat cells. *J. Clin. Endocrinol. Metab.* **76**, 742–747 (1993).
16. Prins, J. B., Walker, N. I., Winterford, C. M. & Cameron, D. P. Apoptosis of human adipocytes *in vitro*. *Biochem. Biophys. Res. Commun.* **201**, 500–507 (1994).
17. Cinti, S. *et al.* Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *J. Lipid Res.* **46**, 2347–2355 (2005).
18. Bhardwaj, R. D. *et al.* Neocortical neurogenesis in humans is restricted to development. *Proc. Natl Acad. Sci. USA* **103**, 12564–12568 (2006).
19. De Vries, H. Atomic bomb effect: variation of radiocarbon in plants, shells, and snails in the past 4 years. *Science* **128**, 250–251 (1958).
20. Nydal, R. & Lovseth, K. Distribution of radiocarbon from nuclear tests. *Nature* **206**, 1029–1031 (1965).
21. Levin, I. & Kromer, B. The tropospheric $^{14}\text{CO}_2$ level in mid latitudes of the northern hemisphere (1959–2003). *Radiocarbon* **46**, 1261–1272 (2004).
22. Spalding, K. L., Buchholz, B. A., Bergman, L. E., Druid, H. & Frisen, J. Forensics: age written in teeth by nuclear tests. *Nature* **437**, 333–334 (2005).
23. Libby, W. F., Berger, R., Mead, J. F., Alexander, G. V. & Ross, J. F. Replacement rates for human tissue from atmospheric radiocarbon. *Science* **146**, 1170–1172 (1964).
24. Harkness, D. D. Further investigations of the transfer of bomb ^{14}C to man. *Nature* **240**, 302–303 (1972).
25. Altman, D. G. *Practical Statistics for Medical Research* pp 164 (Chapman & Hall, CRC, London, 1991).
26. Löfgren, P. *et al.* Long-term prospective and controlled studies demonstrate adipose tissue hypercellularity and relative leptin deficiency in the post-obese state. *J. Clin. Endocrinol. Metab.* **90**, 6207–6213 (2005).
27. Freedman D. S. *et al.* Relationship of childhood overweight to coronary heart disease. Risk factors in adulthood: The Bogalusa Heart Study. *Pediatrics* **108**, 712–718 (2001).
28. Raff, M. C. Size control: the regulation of cell numbers in animal development. *Cell* **26**, 173–175 (1996).
29. Joulia-Ekaza, D. & Cabello, G. Myostatin regulation of muscle development: molecular basis, natural mutations, physiopathological aspects. *Exp. Cell Res.* **312**, 2401–2414 (2006).
30. Stuiver, M. & Polach, H. A. Reporting on ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Stahlberg and T. Bergman for help with high-performance liquid chromatography (HPLC), D. Kurdyla, P. Zermeno and A. Williams for producing graphite, and S. Zdunek for comments on the statistics and modelling. This study was supported by grants from Knut och Alice Wallenbergs Stiftelse, the Human Frontiers Science Program, the Swedish Research Council, the Swedish Cancer Society, the Swedish Heart and Lung foundation, the Novo Nordic Foundation, the Swedish Diabetes Foundation, the Foundation for Strategic Research, the Karolinska Institute, the Tobias Foundation, AFA Life Insurance Health Foundation and NIH/NCRR (RR13461). This work was performed in part under the auspices of the US Department of Energy by University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

Author Contributions K.L.S., P.A. and J.F. designed the study and wrote the manuscript. E.A., P.O.W., S.B., O.B. and T.B. were responsible for the modelling and statistics. K.L.S. and B.A.B. performed sample preparation and ^{14}C accelerator mass spectrometry measurements. L.B., J.H. and E.N. collected clinical material. H.C., M.H. and M.R. performed studies on fat cell purity.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.L.S. (kirsty.spalding@ki.se), J.F. (jonas.frisen@ki.se) or P.A. (peter.arner@ki.se).

LETTERS

The Cl^-/H^+ antiporter CLC-7 is the primary chloride permeation pathway in lysosomes

Austin R. Graves¹, Patricia K. Curran¹, Carolyn L. Smith² & Joseph A. Mindell¹

Lysosomes are the stomachs of the cell—terminal organelles on the endocytic pathway where internalized macromolecules are degraded. Containing a wide range of hydrolytic enzymes, lysosomes depend on maintaining acidic luminal pH values for efficient function. Although acidification is mediated by a V-type proton ATPase, a parallel anion pathway is essential to allow bulk proton transport^{1,2}. The molecular identity of this anion transporter remains unknown. Recent results of knockout experiments raise the possibility that CLC-7, a member of the CLC family of anion channels and transporters, is a contributor to this pathway in an osteoclast lysosome-like compartment, with loss of CLC-7 function causing osteopetrosis³. Several mammalian members of the CLC family have been characterized in detail; some (including CLC-0, CLC-1 and CLC-2) function as Cl^- -conducting ion channels⁴, whereas others act as Cl^-/H^+ antiporters (CLC-4 and CLC-5)^{5,6}. However, previous attempts at heterologous expression of CLC-7 have failed to yield evidence of functional protein, so it is unclear whether CLC-7 has an important function in lysosomal biology, and also whether this protein functions as a Cl^- channel, a Cl^-/H^+ antiporter, or as something else entirely. Here we directly demonstrate an anion transport pathway in lysosomes that has the defining characteristics of a CLC Cl^-/H^+ antiporter and show that this transporter is the predominant route for Cl^- through the lysosomal membrane. Furthermore, knockdown of CLC-7 expression by short interfering RNA can essentially ablate this lysosomal Cl^-/H^+ antiport activity and can strongly diminish the ability of lysosomes to acidify *in vivo*, demonstrating that CLC-7 is a Cl^-/H^+ antiporter, that it constitutes the major Cl^- permeability of lysosomes, and that it is important in lysosomal acidification.

We pursued the function of CLC-7 by using biochemical methods to characterize the Cl^- permeability properties of native lysosomal membranes. We isolated lysosomes from rat liver by using differential sedimentation through a Percoll gradient⁷. This preparation is strongly enriched for LAMP-1, a lysosomal marker, and is strongly depleted for markers of plasma membranes, endosomes, endoplasmic reticulum and mitochondria (Supplementary Fig. 1), suggesting that subsequent functional observations represent transport in lysosomes. We focused on Cl^- transport pathways in these experiments, inhibiting H^+ -ATPases by including no ATP outside the lysosomes.

To assess Cl^- transport, we used the concentrative uptake method^{8,9}. Lysosomes loaded with a high concentration of unlabelled Cl^- were diluted into a buffer containing a trace amount of $^{36}\text{Cl}^-$. If these organelles contain a specific electrogenic transport pathway for Cl^- , they will concentrate the labelled Cl^- inside. Indeed, we observed rapid uptake of $^{36}\text{Cl}^-$, abolished by addition of the K^+ ionophore valinomycin (Fig. 1a, open symbols), indicating a specific electrogenic pathway for the ion (Fig. 1a). (The falling phase of $^{36}\text{Cl}^-$

uptake after 15 s probably represents small leaks of ions through other transport pathways, which dissipate the large transmembrane voltage in these native lysosomes.) Varying internal anions in similar experiments reveals the following apparent permeability sequence: $\text{CH}_3\text{SO}_3^- < \text{I}^- \ll \text{Cl}^- \approx \text{Br}^- < \text{NO}_3^-$ (Fig. 1b). Reduced uptake with I^- inside the lysosomes (Fig. 1b) is reminiscent of the effects of this ion on many CLC family members, which have a low I^- conduction^{10–12}. Uptake was enhanced at pH 4.0 for all permeable ions tested (Fig. 1a, b), which is consistent with activity at physiological lysosomal pH.

pH-dependent concentrative Cl^- uptake could arise either from pH-modulated conduction through an ion channel or from acid activation of a Cl^-/H^+ antiporter. To distinguish between these possibilities we assayed lysosomes for $^{36}\text{Cl}^-$ uptake in symmetrical $[\text{Cl}^-]$ but in the presence of a pH gradient ($[\text{H}^+]_{\text{in}} > [\text{H}^+]_{\text{out}}$). In these conditions, a pH gradient will drive the accumulation of $^{36}\text{Cl}^-$ only if the movement of Cl^- is coupled to that of protons. High levels of $^{36}\text{Cl}^-$ uptake (Fig. 1c, filled symbols), abolished by collapsing the pH gradient with the proton ionophore carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP; open symbols), show the presence of coupled Cl^-/H^+ antiport.

To explore the coupling between Cl^- and H^+ gradients further, we monitored the effects of a Cl^- gradient on the intralysosomal pH by attempting to drive protons uphill with a Cl^- gradient while monitoring internal pH with the ratiometric fluorophore 2',7'-bis-(2-carboxyethyl)-5(6)-carboxyfluorescein (BCECF). The internal pH is stable until valinomycin is added to initiate transport (Fig. 2a, grey arrow). Ensuing alkalinization (red trace) reflects protons being driven out of the lysosomes against their pH gradient, confirming functional Cl^-/H^+ antiport. Subsequent addition of 1 μM FCCP (asterisks) collapses the proton gradient. Control experiments exclude the possibility of a significant proton leak in these conditions (Supplementary Fig. 2).

We distinguished indirect (separate protein pathways for H^+ and Cl^-) from direct coupling mechanisms (obligate antiport of both ions through a common transporter) by measuring the equilibrium potential for H^+ flux, monitoring this flux with BCECF at a series of voltages set with K^+ /valinomycin. Whereas at 0 mV (Fig. 2a, red trace) the internal pH increases on the addition of valinomycin, at –92 mV (dark blue trace, same H^+/Cl^- gradients) the internal pH decreases. Similar measurements at a series of voltages identify the potential at which there is no net H^+ flux (Fig. 2a, green trace, and Fig. 2b). This reversal potential represents thermodynamic equilibrium and can be compared with predictions based on uncoupled H^+ transport (E_{H}) as well as with possible stoichiometries of Cl^-/H^+ antiport calculated with the antiporter equation¹³ (E_{trans}). For the conditions examined in Fig. 2a, an uncoupled H^+ transporter would be predicted to have a reversal potential of +12 mV, the Nernst

¹Membrane Transport Biophysics Unit, and ²Light Microscopy Facility, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 35 Convent Drive, Building 35, MSC 3701, Bethesda, Maryland 20892, USA.

potential for H^+ (E_H , lower pink arrow in Fig. 2b). In contrast, a coupled Cl^-/H^+ antiporter would be predicted to show flux reversal at -54 , -47 or -32 mV (upper pink and red arrows in Fig. 2b) depending on its relative stoichiometry for Cl^- and H^+ . The measured reversal potential is clearly far from that of an uncoupled pathway (Fig. 2b) and agrees well with the predicted reversal potential for a fixed stoichiometry of $2Cl^-:1H^+$. A very different set of ionic conditions (symmetrical Cl^- and a 1-unit pH gradient) similarly yields a reversal potential consistent with a $2Cl^-:1H^+$ stoichiometry (Fig. 2b, blue data and arrows). Thus, the lysosomal transport of Cl^- and H^+ reflects thermodynamically coupled transport through a common protein, a Cl^-/H^+ antiporter with a 2:1 coupling ratio, which provides the major proton pathway in the lysosome (other than the H^+ -ATPase, which is inactive in these experiments).

However, our proton flux measurements do not reveal whether the Cl^-/H^+ antiporter is the major pathway for chloride through lysosomal membranes or is one of several molecular routes for this anion. We can nevertheless assess the contribution of the antiporter to the total Cl^- conductance of lysosomes by measuring the reversal potential for Cl^- flux into these organelles. We monitored lysosomal Cl^- concentration changes with 6-methoxy-*N*-(3-sulphopropyl)quinolinium, inner salt (SPQ), a well-studied fluorescent probe that is effectively quenched by Cl^- (ref. 14). Even though SPQ fluorescence is known to depend somewhat on pH, control experiments (data not shown) reveal the fluorophore to be minimally pH-dependent in the range of ΔpH measured here. We trapped SPQ in lysosomes and measured changes in fluorescence emission on the addition of valinomycin (Fig. 2c). Depending on the membrane voltage (V_m), a single set of Cl^- and H^+ gradients can result in Cl^- efflux ($V_m = -59$ mV; Fig. 2c, upper left), influx ($V_m = -18$ mV or $+20$ mV; Fig. 2c, lower left and lower right) or nearly undetectable flux ($V_m = -25$ mV, Fig. 2c, upper right). By plotting the integral of the difference between initial and final spectra at these voltages, we can estimate the reversal potential for Cl^- flux (Fig. 2d). The measured reversal potential should be the weighted mean of the reversal potentials for each transporter or channel actively conducting Cl^- , reflecting the relative contribution of each permeation pathway to the total flux. The observed flux reverses at about -25 mV, very near the predicted reversal potential of a $2Cl^-:1H^+$ antiporter (Fig. 2d), showing that the vast majority of lysosomal Cl^- flux is through the Cl^-/H^+ antiporter. In combination, the characteristics of this transporter—2:1 Cl^-/H^+ antiport, acid activation, and reduced uptake in I^- —strongly suggest that it is a member of the CLC family.

Which CLC could be the lysosomal Cl^-/H^+ antiporter? Western blots reveal that rat liver expresses the following: CLC-2; at least one of CLC-3, CLC-4 and CLC-5; CLC-6; and CLC-7 (Fig. 3a). Of these CLCs, only CLC-7 is highly enriched in our lysosomes—strongly contrasting

with the others, which were all markedly depleted. Thus, CLC-7 is a prime candidate for the lysosomal antiporter, a function consistent with its role in osteopetrosis^{3,15–17} and in lysosomal storage disease¹⁸.

To establish a molecular connection between CLC-7 and the lysosomal antiporter conclusively, we performed gene-specific knockdown (short interfering RNA (siRNA)) experiments in HeLa cells. These cells express lysosome-localized CLC-7 (Fig. 3a and Supplementary Figs 1 and 4) and possess similar lysosomal Cl^-/H^+ antiport activity to that of rat liver lysosomes (Fig. 3c, black trace, and Supplementary Fig. 5). We knocked down CLC-7 expression with siRNAs directed against two non-overlapping sequences from the gene. With a single transfection of siRNA these cells show somewhat decreased levels of the CLC-7 protein on western blots (Fig. 3b) and show modest, but significant, decreases in H^+ -driven $^{36}Cl^-$ uptake in comparison with untransfected HeLa cells (Fig. 3c; for siRNA 1, $P = 0.0070$; for siRNA2, $P = 0.0092$). Prolonging exposure to the siRNA with two sequential transfections resulted in near-complete knockdown of CLC-7 protein levels (Fig. 3b) as well as near-total loss of H^+ -driven $^{36}Cl^-$ uptake (Fig. 3c, red diamonds; $P = 7 \times 10^{-6}$). Uptake and protein levels were essentially unaffected in HeLa cells transfected with random, scrambled siRNA (Fig. 3a–c). Thus, the CLC-7 protein accounts for essentially all of the observed Cl^-/H^+ antiport activity, which is definitive evidence that CLC-7 is the Cl^-/H^+ antiporter expressed in lysosomes.

If CLC-7 constitutes the long-sought voltage shunt in the lysosomal membrane, disrupting its function would be expected to interfere with lysosomal acidification *in vivo*. We tested this prediction by staining live wild-type (WT) and CLC-7 knockdown HeLa cells with LysoTracker Green, a weakly basic fluorescent probe that stains organelles by virtue of their acidity. We reasoned that if knocking down CLC-7 protein levels affects lysosomal acidification, the knockdown cells should stain less strongly with a LysoTracker dye. Even though LysoTracker may also stain other acidic compartments, any significant disruption in lysosomal acidification should be apparent using this approach. CLC-7 knockdown cells show no gross morphological abnormalities (Fig. 4a, top row). When stained with LysoTracker Green, both WT and control HeLa cells showed dense staining with many punctate objects, presumably lysosomes (Fig. 4a, confocal slice; row 2). In contrast, most CLC-7 knockdown cells showed only a few such puncta, although in every experiment we observed occasional cells with staining comparable to that in the controls (see, for example, Fig. 4a, arrow); these cells may not have been successfully transfected with siRNA, providing a useful control for the imaging conditions. Integrating the total summed intensity from confocal stacks of individual cells (Fig. 4a, row 3) provides a quantitative measure of staining (Fig. 4c) and reveals a significant decrease in CLC-7 knockdown cells compared with either WT ($P = 7.7 \times 10^{-5}$)

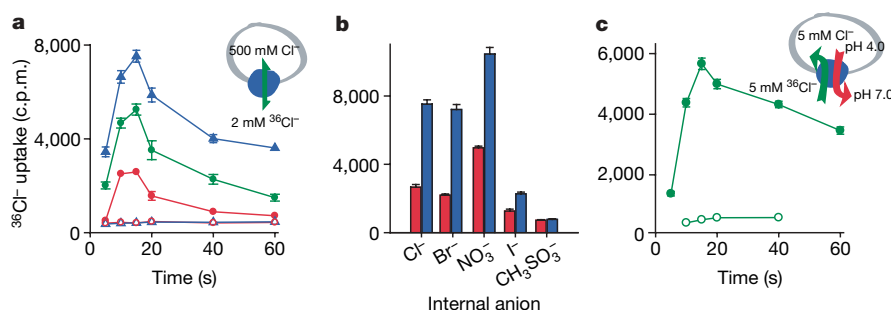


Figure 1 $^{36}Cl^-$ flux in native rat liver lysosomes. **a**, Concentrative uptake by lysosomes. Lysosomes containing 500 mM KCl, pH 7.0 (red), pH 5.5 (green) or pH 4.0 (blue) were added to a solution containing 2 mM $Na^{36}Cl$ at the same pH at $t = 0$ without (filled symbols) or with (open symbols) 1 μM valinomycin. Reactions were stopped at the indicated times and retained radioactivity was measured. **b**, Ion dependence of concentrative uptake. Experiments were performed as above, but with internal Cl^- replaced by the indicated anions; the bars represent $^{36}Cl^-$ uptake at the 20-s time point at

pH 7.0 (red bars) or pH 4.0 (blue bars). **c**, Proton-driven $^{36}Cl^-$ uptake. Experiments were performed as above but with symmetrical Cl^- and a proton driving force (5 mM NaCl at pH 4.0 inside and 5 mM $Na^{36}Cl$ at pH 7.0 with 1 μM valinomycin outside). Data are shown in the absence (closed symbols) or presence (open symbols) of 1 μM FCCP. See Methods for other solution details. $n = 3$ –6; error bars represent s.e.m. and are not shown if they are smaller than the symbols.

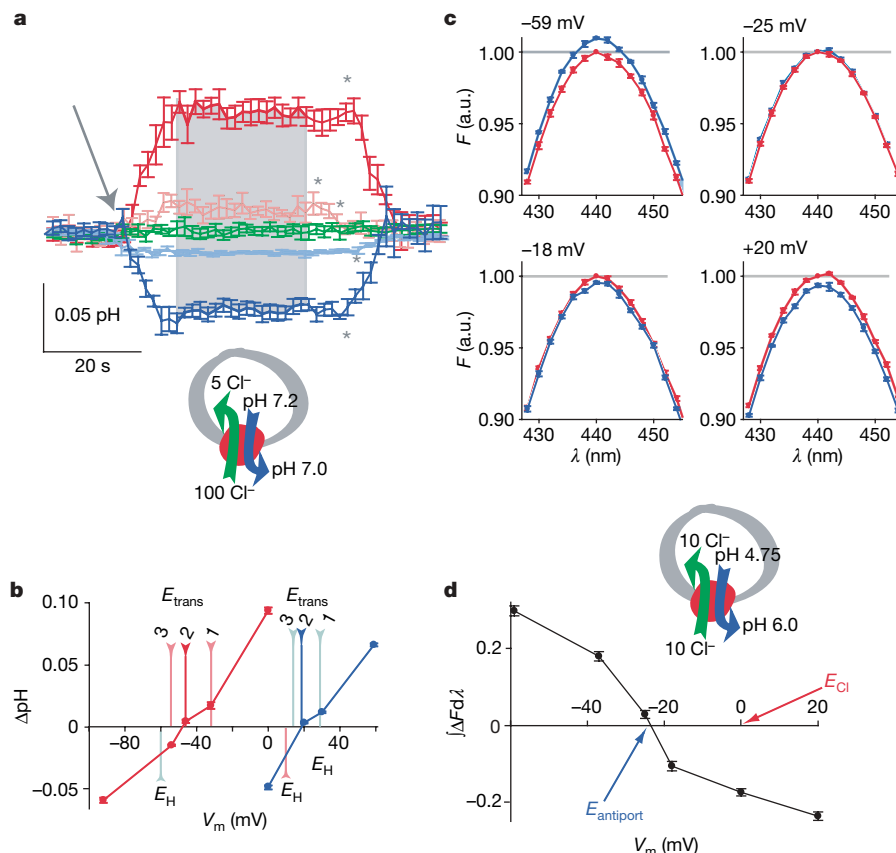


Figure 2 | Fluorescence monitoring of lysosomal H^+ and Cl^- flux. **a**, Cl^- -driven proton movement. ΔpH was measured with the ratiometric BCECF signal (see Methods). $[\text{Cl}^-]_{\text{in}} = 5\text{ mM}$, $\text{pH}_{\text{in}} = 7.2$; $[\text{Cl}^-]_{\text{out}} = 100\text{ mM}$, $\text{pH}_{\text{out}} = 7.0$, $[\text{K}^+]_{\text{out}} = 5\text{ mM}$. Internal K^+ concentration was varied to set E_K at 0 mV (red trace), -32 mV (pink trace), -46 mV (green trace), -56 mV (light blue trace) or -92 mV (dark blue trace). The addition of $1\text{ }\mu\text{M}$ valinomycin initiated transport (arrow). Experiments were terminated by the addition of FCCP (asterisk). $n = 5$ –8; error bars represent s.e.m. **b**, Reversal potentials of Cl^- -driven proton transport. Steady-state ΔpH from two sets of conditions: red symbols represent data from **a**; blue symbols represent $[\text{Cl}^-]_{\text{in}} = [\text{Cl}^-]_{\text{out}} = 10\text{ mM}$, $\text{pH}_{\text{in}} = 6.0$, $\text{pH}_{\text{out}} = 5.0$. Red (pH $7.0/7.2$) and blue (pH $5.0/6.0$) arrows below the x axis indicate predicted

reversal potentials for an uncoupled transporter (E_{H}). Arrows above the x axis indicate predicted reversal potentials for coupled Cl^-/H^+ antiport ($3\text{ Cl}^-:1\text{ H}^+$ (3), $2\text{ Cl}^-:1\text{ H}^+$ (2) or $1\text{ Cl}^-:1\text{ H}^+$ (1)). Each point represents the mean \pm s.e.m. of 13–15 steady-state points (grey area in **a**) from each of five to eight experiments. **c**, Baseline emission spectra at indicated voltages (red symbols, normalized to 1.00 at 440 nm, grey line) with 10 mM Cl^- , $\text{pH } 4.75$ inside lysosomes and 10 mM Cl^- , $\text{pH } 6.0$ outside; the addition of $1\text{ }\mu\text{M}$ valinomycin initiated Cl^- flux; after 2 min a second spectrum was obtained (blue symbols). **d**, Reversal potential for Cl^- flux: each point represents the integrated difference spectrum at the given voltage (area between spectra in **c**). The arrows indicate E_{Cl} (red) and $E_{2:1}$ (blue, the predicted reversal potential for a $2:1\text{ Cl}^-/\text{H}^+$ antiporter). $n = 4$ –6; error bars represent s.e.m.

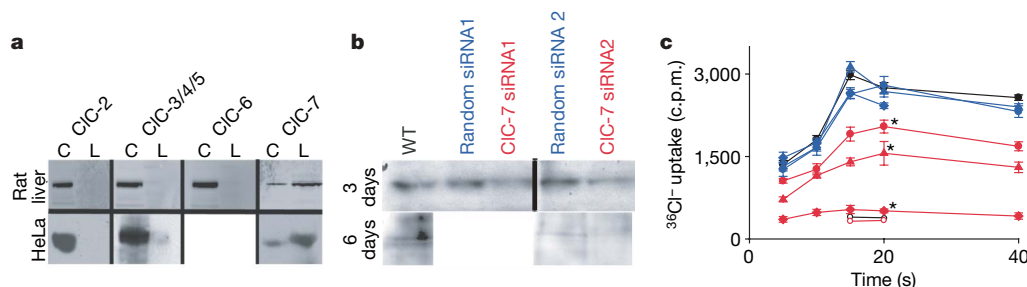


Figure 3 | CIC-7 mediates lysosomal Cl^-/H^+ antiport. **a**, Western blots of CLCs in rat liver and HeLa cells. For each CLC shown, blots indicate expression levels in crude lysate (C) and the lysosome-enriched fraction (L). The antibody against the CIC-3/4/5 subfamily cross-reacts with all three members of that group. No CIC-6 expression was observed in HeLa cells. **b**, Western blots of CIC-7 in whole-cell lysates of HeLa cells either untransfected (WT) or transfected with either random non-coding siRNA or with one of two different CIC-7-specific siRNAs. Cells were transfected with either a three-day single-stage protocol or a six-day two-stage protocol (see Methods for details). The conditions used for the three-day and six-day blots were different; WT controls are included in both cases as references for baseline CIC-7 expression. **c**, H^+ -driven $^{36}\text{Cl}^-$ uptake in lysosomes isolated

from WT HeLa cells (filled black symbols), three different preparations of HeLa cells transfected with control siRNA (filled blue symbols; circles and triangles represent single transfection, diamonds represent double transfection) or HeLa cells transfected with two different CIC-7-specific siRNAs (filled red triangles, filled red circles and filled red diamonds). For one of the CIC-7 siRNAs, data are shown for both a single siRNA transfection (red triangles) and a two-stage transfection (red diamonds; see Methods for details). FCCP controls are shown at two time points for WT (open black symbols) and one CIC-7 siRNA (open red symbols). Asterisks indicate statistically significant differences compared with WT (siRNA1, 3 days: $P = 0.0092$; siRNA2, 3 days: $P = 0.0070$; siRNA2, 6 days: $P = 7 \times 10^{-6}$; two-tailed Student's t -test). $n = 3$; error bars represent s.e.m.

or control-siRNA-transfected ($P = 3.3 \times 10^{-9}$) HeLa cells. (Note that the control-siRNA-transfected cells show a marginal increase in LysoTracker staining.) When stained with an anti-LAMP-1 antibody, WT, siRNA control and CIC-7 siRNA cells showed similar staining patterns (Fig. 4b), suggesting that lysosomes still exist in the CIC-7 knockdown cells. In combination with the results above, which point to a specific lysosomal effect of CIC-7 knockdown, these results demonstrate that a decrease in CIC-7 *in vivo* compromises lysosomal acidification and support our conclusion that this transporter represents the major pathway for Cl^- in this organelle.

We have established that CIC-7 is a Cl^-/H^+ antiporter, revealing that the last remaining uncharacterized subfamily of CLCs consists of antiporters rather than ion channels. This leads to the general conclusion that for CLCs, subcellular localization correlates strictly with function; the two subfamilies of intracellular CLCs are both proton-coupled Cl^- transporters, whereas the plasma-membrane CLCs are all ion channels. Our results establish the stoichiometry of CIC-7 as

$2\text{Cl}^-:1\text{H}^+$; they are in agreement with the approximate measurements already made for CIC-4 and CIC-5 (refs 5, 6) and with the measurement for the plant antiporter AtClCa (ref. 19), supporting the contention that the basic mechanism of antiport is conserved between bacterial, plant and mammalian CLC transporters.

The present results have functional implications for the cell biology of the lysosome, because CIC-7 accounts for the major anion pathway of this organelle. Previous results have shown that lysosomal acidification depends on the presence of external Cl^- (refs 1, 2); our knockdown experiments reveal that this pathway is essential for the lysosomal acidification mechanism. Although the existence of such a pathway has long been inferred¹, its molecular identity is now clear. A series of CLC antiporters participate in acidification in the endosomal pathway, with CIC-4 and CIC-5 involved in early endosomes, CIC-6 localizing to late endosomes, and CIC-7 involved in lysosomes²⁰. These compartments have progressively more acidic interiors; because they express the same H^+ -ATPase, we speculate that cells vary the identity of the CLC transporter to help in determining the final pH setpoint of each organelle in the endocytic pathway. Because a loss of CIC-7 function leads to osteopetrosis, it has been suggested that a CIC-7 inhibitor could serve as a treatment for osteoporosis²¹; the functional assay system described here could be of use in high-throughput screening for such inhibitors.

METHODS SUMMARY

Lysosomal enrichment from rat liver and HeLa cell culture. Saline-perfused rat livers were homogenized with a motorized Potter–Elvehjem homogenizer. Lysosomes were isolated by differential centrifugation on a Percoll gradient⁷ and fractions were identified with enzyme assays or western blots with anti-LAMP-1 antibody.

Concentrative and proton-driven $^{36}\text{Cl}^-$ uptake. Lysosomal samples containing 100 μg of total protein were loaded by freeze–thawing followed by sonication in the desired buffer, and external solutions were exchanged by using Sephadex G-50 columns equilibrated in external buffer. Concentrative uptake was initiated by adding 2 mM $^{36}\text{Cl}^-$ to the outside buffer; proton-driven $^{36}\text{Cl}^-$ uptake was initiated by the addition of 1 μM valinomycin to the outside buffer. Reactions were terminated by filter binding and radioactivity was measured by liquid scintillation.

Fluorescence measurements of H^+ and Cl^- flux. BCECF or SPQ (200 μM) was trapped in lysosomes along with internal solution by freeze–thawing followed by sonication; external solutions were replaced as above. Lysosomes were introduced to a Jobin–Yvon Fluoromax-3 fluorimeter and stirred at 22 °C. BCECF excitation wavelengths alternated between 500 and 450 nm; emission was measured at 535 nm. SPQ was excited at 344 nm; emission spectra (420–460 nm) were collected. Transport was initiated by adding 1 μM valinomycin.

Cell culture and siRNA-mediated knockdown of CIC-7. HeLa cells were grown to generate a harvest of about 6 g of cells for lysosome preparations or on chambered coverglasses for imaging. Cells were transfected with 46 nM siRNA complexed with siPORT NeoFx (Ambion) and either collected at 72 h (single transfection) or re-transfected for a further 72 h (double transfection). Collected cells were lysed by using nitrogen cavitation²², and lysosomes were prepared as above. The siRNA sequences used were as follows: CIC-7 siRNA1, GGCCUCAUCAUCUGGAAtt (104370; Ambion); CIC-7 siRNA2, CCUCUCCGAGUUGAUAACtt (145731; Ambion).

Imaging of HeLa cells. Cells were imaged with a Zeiss LSM 510 confocal microscope after 10 min in 50 nM LysoTracker Green (Invitrogen). Similar results were obtained for at least four transfections. For immunostaining, fixed cells were stained with anti-LAMP-1 primary antibody and an Alexa-546 conjugated secondary antibody.

Received 28 September 2007; accepted 10 March 2008.
Published online 30 April 2008.

1. Dell'Antone, P. Evidence for an ATP-driven 'proton pump' in rat liver lysosomes by basic dyes uptake. *Biochem. Biophys. Res. Commun.* **86**, 180–189 (1979).
2. Ohkuma, S., Moriyama, Y. & Takano, T. Identification and characterization of a proton pump on lysosomes by fluorescein-isothiocyanate-dextran fluorescence. *Proc. Natl Acad. Sci. USA* **79**, 2758–2762 (1982).
3. Kornak, U. *et al.* Loss of the CIC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* **104**, 205–215 (2001).
4. White, M. M. & Miller, C. A voltage-gated anion channel from the electric organ of *Torpedo californica*. *J. Biol. Chem.* **254**, 10161–10166 (1979).

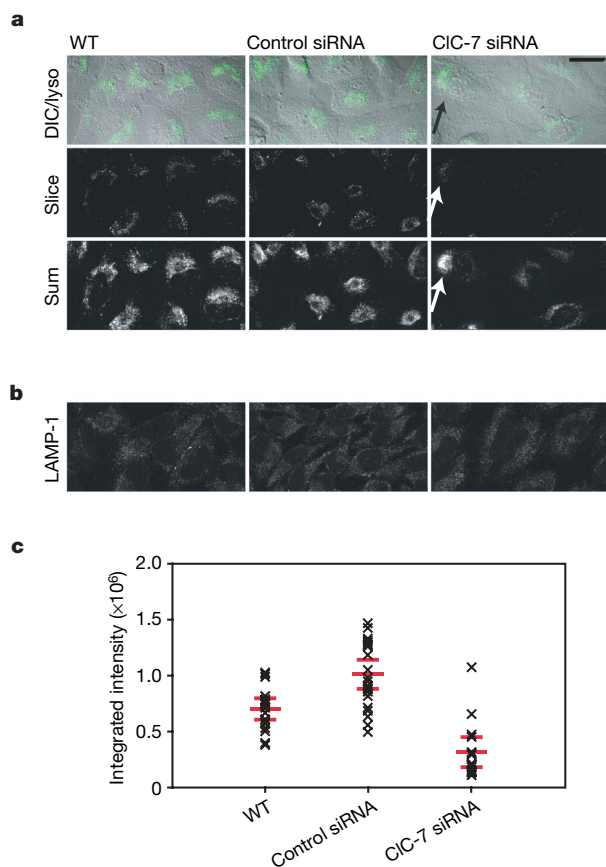


Figure 4 | CIC-7 is essential for lysosomal acidification *in vivo*. **a**, Confocal images of live WT, control-siRNA-transfected and CIC-7-siRNA-transfected HeLa cells stained with LysoTracker Green dye. Top row, differential interference contrast images overlaid with summed LysoTracker intensity (green); middle row, the brightest single confocal slice from the LysoTracker Green channel for the same field; bottom row, summed LysoTracker Green intensity for all confocal slices through the total thickness of the field. The arrows indicate a cell in the CIC-7 knockdown condition with a LysoTracker staining level roughly the same as that of the WT. **b**, Confocal images of WT, control-siRNA-transfected and CIC-7-siRNA-transfected HeLa cells fixed and stained with anti-LAMP-1 antibody. **c**, Integrated LysoTracker Green staining intensity of summed slices through WT, control-siRNA-transfected and CIC-7-siRNA-transfected HeLa cells. The intensity for each measured cell is denoted by a cross, with the mean being indicated by a long red bar; s.e.m. is denoted by short red bars. Differences between CIC-7-siRNA-transfected and WT cells and between CIC-7-siRNA-transfected and control-siRNA-transfected cells were significant ($P = 7.7 \times 10^{-5}$ and 3.3×10^{-9} , respectively). Scale bar, 29 μm . Each row of images in **a** and **b** was collected with identical imaging settings and was processed identically.

5. Picollo, A. & Pusch, M. Chloride/proton antiporter activity of mammalian CLC proteins CIC-4 and CIC-5. *Nature* **436**, 420–423 (2005).
6. Scheel, O., Zdebik, A. A., Lourdel, S. & Jentsch, T. J. Voltage-dependent electrogenic chloride/proton exchange by endosomal CLC proteins. *Nature* **436**, 424–427 (2005).
7. Graham, J. M. in *Current Protocols in Cell Biology* (eds Bonifacino, J. S., Dasso, M., Harford, J. B., Lippincott-Schwartz, J. & Yamada, K. M.) 3.6.1–3.6.21 (Wiley, Hoboken, NJ, 2000).
8. Goldberg, A. F. & Miller, C. Solubilization and functional reconstitution of a chloride channel from *Torpedo californica* electroplax. *J. Membr. Biol.* **124**, 199–206 (1991).
9. Garty, H., Rudy, B. & Karlish, S. J. A simple and sensitive procedure for measuring isotope fluxes through ion-specific channels in heterogenous populations of membrane vesicles. *J. Biol. Chem.* **258**, 13094–13099 (1983).
10. Rychkov, G. Y., Pusch, M., Roberts, M. L., Jentsch, T. J. & Bretag, A. H. Permeation and block of the skeletal muscle chloride channel, CIC-1, by foreign anions. *J. Gen. Physiol.* **111**, 653–665 (1998).
11. Jentsch, T. J., Friedrich, T., Schriever, A. & Yamada, H. The CLC chloride channel family. *Pflügers Arch.* **437**, 783–795 (1999).
12. Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic CIC-type chloride channel. *J. Gen. Physiol.* **114**, 713–722 (1999).
13. Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of CIC Cl[−] channels. *Nature* **427**, 803–807 (2004).
14. Illsley, N. P. & Verkman, A. S. Membrane chloride transport measured using a chloride-sensitive fluorescent probe. *Biochemistry* **26**, 1215–1219 (1987).
15. Henriksen, K. *et al.* Characterization of osteoclasts from patients harboring a G215R mutation in CIC-7 causing autosomal dominant osteopetrosis type II. *Am. J. Pathol.* **164**, 1537–1545 (2004).
16. Letizia, C. *et al.* Type II benign osteopetrosis (Albers-Schonberg disease) caused by a novel mutation in CLCN7 presenting with unusual clinical manifestations. *Calcif. Tissue Int.* **74**, 42–46 (2004).
17. Kornak, U., Ostertag, A., Branger, S., Benichou, O. & de Vernejoul, M. C. Polymorphisms in the CLCN7 gene modulate bone density in postmenopausal women and in patients with autosomal dominant osteopetrosis type II. *J. Clin. Endocrinol. Metab.* **91**, 995–1000 (2006).
18. Kasper, D. *et al.* Loss of the chloride channel CIC-7 leads to lysosomal storage disease and neurodegeneration. *EMBO J.* **24**, 1079–1091 (2005).
19. De Angeli, A. *et al.* The nitrate/proton antiporter AtCLCa mediates nitrate accumulation in plant vacuoles. *Nature* **442**, 939–942 (2006).
20. Jentsch, T. J. Chloride and the endosomal–lysosomal pathway: emerging roles of CLC chloride transporters. *J. Physiol. (Lond.)* **578**, 633–640 (2007).
21. Schaller, S. *et al.* The chloride channel inhibitor NS3736 prevents bone resorption in ovariectomized rats without changing bone formation. *J. Bone Miner. Res.* **19**, 1144–1153 (2004).
22. Gottlieb, R. A. & Adachi, S. Nitrogen cavitation for cell disruption to obtain mitochondria from cultured cells. *Methods Enzymol.* **322**, 213–221 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Swartz, M. Maduke, J. Diamond and R. Youle for critical readings of the manuscript; R. Brady, G. Murray and R. Puertollano-Moro for advice on lysosomes; and the members of the Mindell laboratory for discussions. This work was supported by the NINDS intramural program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.M. (mindellj@ninds.nih.gov).

Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells

Yoshinori Kawamura^{1*}, Kuniaki Saito^{1,2*}, Taishin Kin³, Yukiteru Ono⁴, Kiyoshi Asai^{3,5}, Takafumi Sunohara¹, Tomoko N. Okada^{1,2}, Mikiko C. Siomi^{1,2,6} & Haruhiko Siomi^{1,2}

RNA silencing is a conserved mechanism in which small RNAs trigger various forms of sequence-specific gene silencing by guiding Argonaute complexes to target RNAs by means of base pairing^{1,2}. RNA silencing is thought to have evolved as a form of nucleic-acid-based immunity to inactivate viruses and transposable elements. Although the activity of transposable elements in animals has been thought largely to be restricted to the germ line, recent studies have shown that they may also actively transpose in somatic cells, creating somatic mosaicism in animals³. In the *Drosophila* germ line, Piwi-interacting RNAs arise from repetitive intergenic elements including retrotransposons by a Dicer-independent pathway and function through the Piwi subfamily of Argonautes to ensure silencing of retrotransposons^{4–9}. Here we show that, in cultured *Drosophila* S2 cells, Argonaute 2 (AGO2), an AGO subfamily member of Argonautes, associates with endogenous small RNAs of 20–22 nucleotides in length, which we have collectively named endogenous short interfering RNAs (esiRNAs). esiRNAs can be divided into two groups: one that mainly corresponds to a subset of retrotransposons, and the other that arises from stem-loop structures. esiRNAs are produced in a Dicer-2-dependent manner from distinctive genomic loci, are modified at their 3' ends and can direct AGO2 to cleave target RNAs. Mutations in *Dicer-2* caused an increase in retrotransposon transcripts. Together, our findings indicate that different types of small RNAs and Argonautes are used to repress retrotransposons in germline and somatic cells in *Drosophila*.

Biochemical and mutation analyses have revealed the existence of multigene families encoding two key proteins—Dicer and Argonaute—in RNA silencing in *Drosophila*^{10–13}. Dicer-1 generates *Drosophila* microRNAs (miRNAs) whereas Dicer-2 creates short

interfering RNAs (siRNAs). Argonaute proteins directly bind small guide RNAs and either display endonucleolytic ('Slicer') activity or have a platform role for the assembly of silencing complexes². The five *Drosophila* Argonaute proteins can be subdivided into two subfamilies: the ubiquitous AGO (AGO1 and AGO2) and the germline-specific Piwi (AGO3, Aubergine (Aub) and Piwi) subfamilies^{2,14}. AGO1 is involved in the miRNA-dependent pathway that silences messenger RNA, whereas AGO2 functions in RNA interference (RNAi)¹² directed by exogenous siRNAs. Piwi proteins are involved in silencing retrotransposons through direct interaction with Piwi-interacting RNAs (piRNAs) in the germ line^{4–9} by forming a Dicer-independent cycle that amplifies piRNAs^{6,7}. In contrast, how RNA silencing might operate to repress retrotransposons in *Drosophila* somatic cells, and to what extent, remain unknown.

Unlike for the other Argonaute proteins in *Drosophila*, endogenous small RNA partners for AGO2 have not been identified. AGO2 could exist simply to lie in wait for small RNAs that exogenously come into contact with it, including those originating from infecting viruses that produce double-stranded RNAs (dsRNAs) as part of their replication cycle¹⁵. However, we surmised that AGO2 might also interact with endogenous siRNAs that are products of Dicer processing of dsRNAs, akin to endogenous siRNAs in *Arabidopsis thaliana*¹. In animals, endogenous siRNAs have only been identified in *Caenorhabditis elegans*¹⁶. To test whether AGO2 exists in a complex with endogenous siRNAs, we immunopurified AGO2 with a specific monoclonal antibody from a cultured *Drosophila* somatic S2 cell line (Supplementary Fig. 1a, b) and examined its associated RNAs. AGO2 in S2 cells was predominantly associated with small RNAs of about 21 nucleotides in length (Supplementary Fig. 1c).

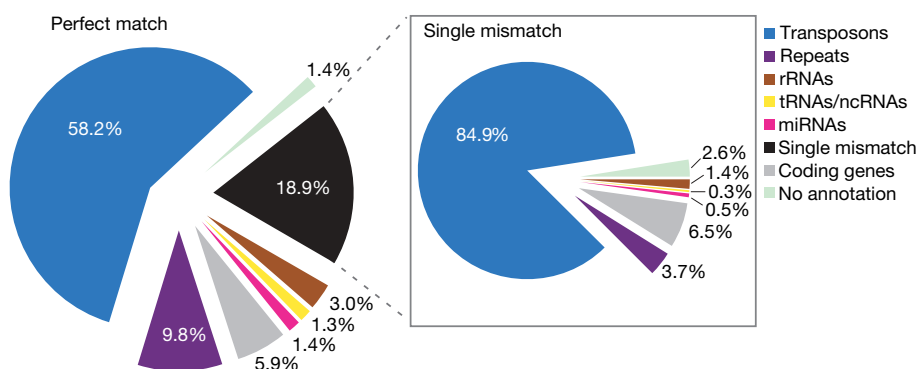


Figure 1 | Identification of endogenous small RNAs that bind AGO2. The left chart shows that the contents of AGO2-associated small RNAs perfectly match the *Drosophila* genome sequence (perfect match). Clones with single mismatches comprise 12,240 from a total of 64,588; these populations are shown on the right (single mismatch).

¹Institute for Genome Research, University of Tokushima, Tokushima 770-8503, Japan. ²Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. ³Computational Biology Research Center, Advanced Industrial Science and Technology, Tokyo 135-0664, Japan. ⁴Information and Mathematical Science Laboratory, Inc., Tokyo 112-0012, Japan. ⁵Graduate School of Frontier Science, University of Tokyo, Chiba 277-8561, Japan. ⁶Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST), Saitama 332-0012, Japan.

*These authors contributed equally to this work.

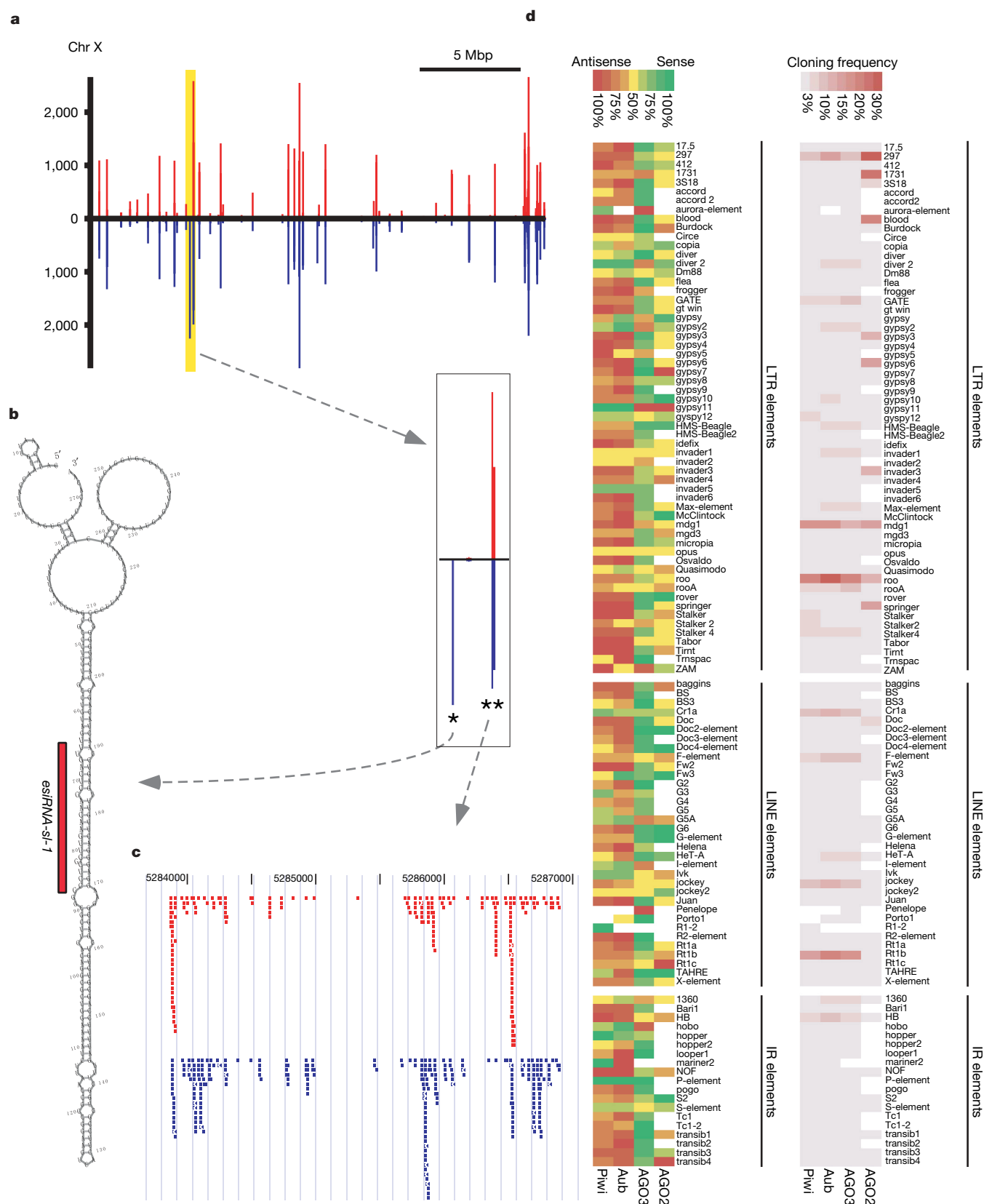


Figure 2 | esiRNA clusters. **a**, Frequency map of AGO2-associated small RNAs on chromosome X. The x and y axes represent the chromosomal position and the number of reads, respectively. The strand is represented by the colour of the bars: plus (red) and minus (blue). Two major hotspots indicated by the yellow box are enlarged in the rectangle below; one, indicated with the asterisk, expresses only from the minus strand, where 20 similar sequence patterns (~274 nt) are repeatedly encoded (Supplementary Fig. 5). Each fragment has a distinct stem-loop structure, as shown in **b**. The most abundant small RNA (*esiRNA-si-1*) derived from this fragment is indicated by a red bar. The other hotspot (indicated with two asterisks) expresses esiRNAs

from both strands, with details shown in **c**. **d**, The heat map (left panel) indicates the strand bias of cloned esiRNAs and piRNAs⁷ with respect to canonical transposon sequences (indicated at the right). Transposons are grouped into long terminal repeat (LTR), long interspersed nuclear (LINE) and inverted repeat (IR) elements. The colour intensities indicate the degree of strand bias (green, sense; red, antisense; yellow, unbiased). Note that some transposons in the AGO2 lane are coloured either green or red, but the cloning frequency of these transposons was very low (see the right panel). The cloning frequencies of individual transposons in all four complexes (this study and ref. 7) are indicated as a heat map (right panel).

To characterize the small RNAs associated with AGO2, we constructed a complementary DNA library and performed high-throughput pyrosequencing¹⁷ that generated 77,327 reads of the cDNA library. Consistent with the [γ -³²P]ATP-labelled result, the sizes of the small RNAs ranged from 20 to 22 nucleotides in length (Supplementary Fig. 1d), similar to the sizes of miRNAs but clearly different from the sizes of piRNAs (24–29 nucleotides). Most of these endogenous binding partners for AGO2 were a previously undescribed class of ~21-nucleotide small RNAs (see below), which we have collectively referred to as esiRNAs (also known as endo-siRNAs). Unlike many other small RNAs that bind Argonautes^{1,18}, esiRNAs do not display a nucleotide-bias at any particular position (data not shown). Among the 77,327 reads, 52,348 sequences matched the *Drosophila* genome with 100% identity over their entire length (Fig. 1). In addition, a large number of the AGO2-associated small RNAs showed single mismatches (12,168; Fig. 1). Adenosine-to-guanosine single mismatches were vastly overrepresented (Supplementary Fig. 1e), suggesting that a portion of the small RNAs may be adenosine-to-inosine RNA edited by ADAR (adenosine deaminase acting on RNA) enzymes¹⁹.

Database searching revealed that a large number of the AGO2-associated small RNAs corresponded to transposons and other repetitive elements in the genome (Fig. 1). Pieces of protein-coding sense mRNAs (~6%) as well as ribosomal RNAs (~3%) were conspicuous in this screening, as is the case for piRNAs in fly⁹ and mouse²⁰. We also observed a small number of miRNAs (~1.4%) among the AGO2-associated small RNAs sequenced (Supplementary Table 1), including two previously unknown miRNAs (*miR-20071* and *miR-20072*) derived from introns of *prp8* and *cdc2*, respectively (Supplementary Figs 2 and 3).

We summed the number of unique small RNAs in a 5-kilobase (kb) sliding window and plotted this against the *Drosophila* draft assembly (Berkeley *Drosophila* Genome Project release 5). Clusters of small RNA production were observed (Fig. 2a, Supplementary Fig. 4 and Supplementary Table 2). Most esiRNAs could be divided into two types: those that matched retrotransposons and those that arose from long stem-loop structures from repetitive sequences located in an intergenic region on chromosome X (Fig. 2b, c and Supplementary Fig. 5). Populations of esiRNAs derived from retrotransposons are quite complex, with most RNAs being cloned only once (57.2% in perfect and single matches). Like *Drosophila* piRNAs^{4–9}, esiRNAs are derived from a subset of retrotransposons (Fig. 2d). Consistent with this, a number of esiRNAs (24.5%) overlap with registered piRNA sequences^{6–9}; however, their sizes (20–22 nucleotides versus 24–27 nucleotides) and binding partners (AGO2 versus Piwi proteins) were

clearly different. Thus, esiRNAs and piRNAs are distinct classes of endogenous small RNAs. Some retrotransposons including 1731 and 297 elements are overrepresented in the esiRNA populations. *Drosophila* piRNAs show a strong strand bias for sequences; Aub and Piwi mainly bind antisense-strand piRNAs, whereas AGO3 binds sense-strand piRNAs^{6,7}. esiRNAs, however, seems to be unbiased (Fig. 2d and Supplementary Fig. 4). For example, one hotspot for small RNA production within a 7,600-nucleotide intergenic region on chromosome X was almost entirely covered by small RNAs on both strands (Fig. 2c). Thus, they must be derived from dsRNAs.

In contrast, esiRNAs arising from stem-loop structures (Fig. 2b) showed strand bias with exceptionally high cloning frequency (Supplementary Fig. 5). The stem-loop-generating region contains twenty ~274-nucleotide repeats and appeared to extend to the CG4068 gene (Supplementary Fig. 5). All of the esiRNAs from this region were found in a sense orientation with respect to the orientation of CG4068 (Supplementary Fig. 5). Northern blotting revealed that both types of esiRNAs specifically co-purified with AGO2, but not with AGO1 (Supplementary Fig. 6a–c). An *in vitro* Slicer assay also confirmed that AGO2–esiRNA complexes were capable of specifically cleaving target RNAs containing sequences perfectly complementary to esiRNAs (Fig. 3a). This also implied that esiRNAs can guide AGO2 to cleave complementary RNAs including retrotransposon transcripts in cells.

piRNAs and exogenous siRNAs in *Drosophila* are modified with 2'-O-methylation at their 3' terminus, requiring an RNA methyltransferase, Pimet (also known as DmHen1 and CG12367) (refs 21 and 22). esiRNAs in S2 cells were resistant to periodate oxidation (NaIO₄) and β -elimination reactions (Fig. 3b, c), but esiRNAs from *DmHen1/Pimet* mutant ovaries could be β -eliminated (Fig. 3d). Thus, esiRNAs are likely to have the same 3' modification, as is the case for piRNAs and exogenous siRNAs. These findings suggested that esiRNAs are produced by a pathway similar to that of exogenous siRNAs that bind AGO2 in S2 cells. Because exogenous siRNA could not displace esiRNAs from AGO2 (Supplementary Fig. 7a, b), 'empty' AGO2 (apo-AGO2) seems to exit to accommodate small RNAs from exogenous sources.

To test whether the production and normal accumulation of esiRNAs require Dicers and other known small-RNA-processing factors, we depleted each of these proteins from S2 cells by RNAi and monitored the abundance of small RNAs (Fig. 4a). A marked reduction of *esiRNA-sl-1*, one of the very abundant esiRNAs (Fig. 2b), was observed in Dicer-2-depleted S2 cells; this was observed less efficiently in AGO2-depleted S2 cells, and was not observed in Drosha-, Dicer-1- or AGO1-depleted S2 cells. We also confirmed

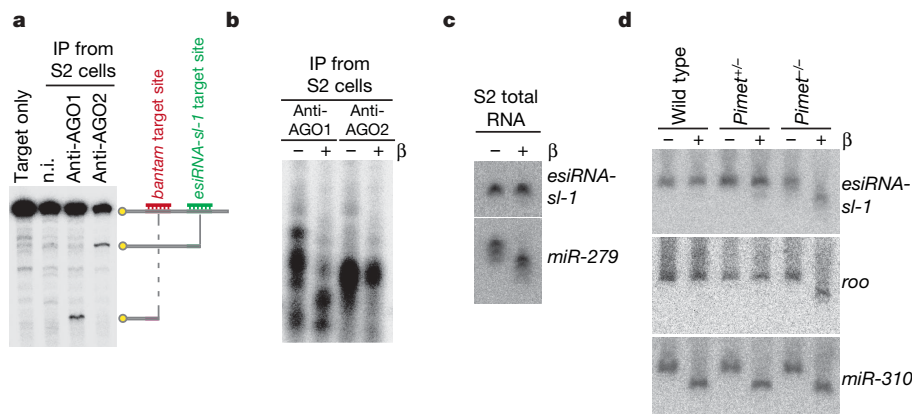


Figure 3 | Specific association and modification of esiRNAs. **a**, Target RNAs containing sequences complementary to *bantam* (red bar) and *esiRNA-sl-1* (green bar) are cleaved at each position with AGO1–miRNA and AGO2–esiRNA complexes, as expected. n.i.; non-immune IgG; IP, immunoprecipitation. **b**, Unlike miRNAs, esiRNAs show resistance to

β -elimination. **c**, Total RNAs of S2 cells were probed for *esiRNA-sl-1* and *miR-279* before and after β -elimination. **d**, Loss of *Pimet* causes esiRNA to be sensitive to β -elimination, suggesting that esiRNAs are methylated by Pimet. *roo*, piRNA corresponding to the LTR retrotransposon *roo*.

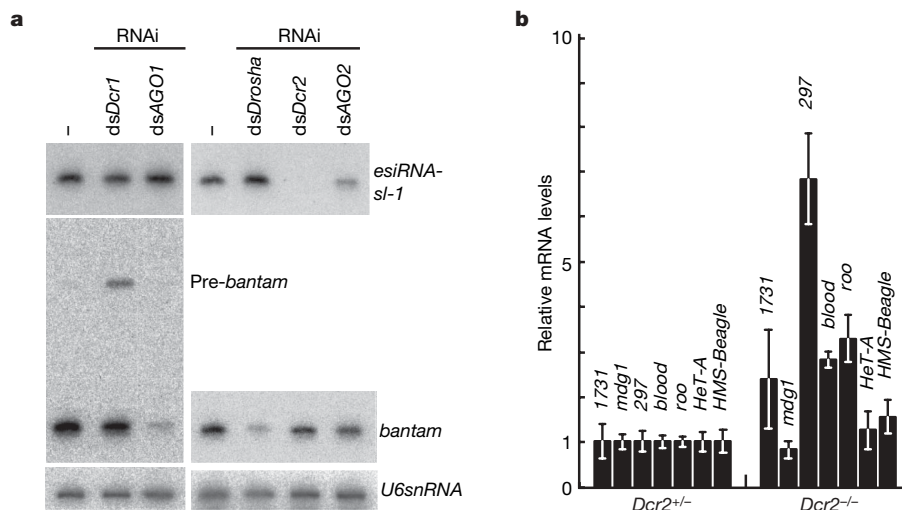


Figure 4 | esiRNAs are produced in a Dicer-2-dependent manner, and are involved in transposon repression in somatic cells. a, Dicer-2 (*Dcr2*) depletion causes a severe reduction of esiRNAs. **b**, Increase of retrotransposon transcript abundance in homozygous *Dcr2* mutants. The steady-state levels of transcripts from retrotransposons were measured by

quantitative RT-PCR in carcasses (flies without testes) of heterozygous or homozygous *Dcr2* mutant males, relative to *rp49* (also known as *RpL32*). The error bars report the average \pm s.d. for at least three independent experiments.

that esiRNAs were almost undetectable in the ovaries of *Dicer-2* mutants and were significantly reduced in *AGO2* mutants (Supplementary Fig. 8). Although the normal accumulation of piRNAs requires *Spindle-E* (*Spn-E*)⁴, a lack of the *Spn-E* activity does not affect the levels of esiRNAs in the ovary (Supplementary Fig. 8). These results demonstrated that esiRNAs are produced in a Dicer-2-dependent manner.

We examined whether the somatic expression of retrotransposons is suppressed by the esiRNA pathway. Depletion of Dicer-2 or AGO2 coincided with higher levels of retrotransposon transcripts (Supplementary Fig. 9a, b). Mutations in *Dicer-2* also caused an increase in retrotransposon transcripts (Fig. 4b). We therefore concluded that the expression of a subset of retrotransposons was reduced by the esiRNA pathway in somatic cells.

Transposable elements are powerful mutator elements responsible for generating variations in the host genome, and therefore have a role as crucial factors shaping the genome through evolution²³. Previous studies have suggested that expression of transposable elements in animal somatic cells is mainly silenced by DNA methylation at transcriptional levels²³. However, we found that the expression of a subset of retrotransposons was reduced by the esiRNA pathway that requires Dicer-2 and AGO2 in *Drosophila* somatic cells. This was clearly distinct from the silencing of retrotransposons by the piRNA pathway that requires Piwi proteins but not Dicers in the germ line^{4,6,7}. Our findings, in turn, suggested that some retrotransposons are actively transcribed in somatic cells, which are then processed into small RNAs by Dicer-2. Recent studies in the fission yeast *Schizosaccharomyces pombe* have shown that RNAi-mediated heterochromatin formation requires the production of RNA from heterochromatic regions to generate small RNAs and serve as part of a feed-forward loop required to maintain the heterochromatic state²⁴. Thus, it is tempting to speculate that esiRNAs could not only guide AGO2 to cleave transcripts of retrotransposons but are also able to direct silencing complexes to specific chromosomal regions. Interestingly, a loss of *Dicer-2* or *AGO2* activity is correlated with defects in the formation of heterochromatin in *Drosophila*^{25,26}. These results together suggest that changes in the activity of Dicer-2 and/or AGO2 could allow the expression or even transposition of retrotransposons that may in turn create quantitative genetic variation in gene expression within somatic cell populations.

METHODS SUMMARY

AGO1 and AGO2 were immunopurified from S2 cells using specific antibodies²⁷. Cloning of small RNAs associated with AGO2 in S2 cells was carried out as described⁹. Genome mapping and annotation was performed as described in the Methods section. Northern blotting⁵, the *in vitro* target-RNA cleavage assay⁹, RNAi for S2 cells¹², peroxidation/ β -elimination^{6,21} and RT-PCR analysis⁴ were performed as described.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2007; accepted 2 April 2008.

Published online 7 May 2008.

- Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nature Rev. Genet.* **8**, 884–896 (2007).
- Peters, L. & Meister, G. Argonaute proteins: mediators of RNA silencing. *Mol. Cell* **26**, 611–623 (2007).
- Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
- Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
- Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20**, 2214–2222 (2006).
- Gunawardane, L. S. *et al.* A Slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
- Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
- Yin, H. & Lin, H. An epigenetic activation role of Piwi and a Piwi-associated piRNAs in *Drosophila melanogaster*. *Nature* **450**, 304–308 (2007).
- Nishida, K. M. *et al.* Gene silencing mechanisms mediated by Aubergine-piRNA complexes in *Drosophila* male gonad. *RNA* **13**, 1911–1922 (2007).
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
- Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
- Williams, R. W. & Rubin, G. M. ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc. Natl Acad. Sci. USA* **99**, 6889–6894 (2002).
- Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452–454 (2006).

16. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
17. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
18. Kim, V. N. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev.* **20**, 1993–1997 (2006).
19. Nishikura, K. Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nature Rev. Mol. Cell Biol.* **7**, 919–931 (2006).
20. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
21. Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* **21**, 1603–1608 (2007).
22. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
23. Muotri, A. R., Marchetto, M. C. N., Coufal, N. G. & Gage, F. H. The necessary junk: new functions for transposable elements. *Hum. Mol. Genet.* **16**, R159–R167 (2007).
24. Grewal, S. I. & Elgin, S. C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406 (2007).
25. Peng, J. C. & Karpen, G. H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell Biol.* **9**, 25–35 (2007).
26. Deshpande, G., Calhoun, G. & Schedl, P. *Drosophila* argonute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev.* **19**, 1680–1685 (2005).
27. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M. C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to T. Suzuki and members of the Siomi laboratory for discussion and comments on this manuscript. We thank K. Yamada and E. Hattori for expert assistance in AGO2-associated small RNA mapping and annotation, N. Iwanami and T. Hirose for help in qRT-PCR, and M. Itakura for continuous support and encouragement. This work was supported by MEXT grants to M.C.S. and H.S., a MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) 21st COE (Centers of Excellence) postdoctoral fellowship to K.S. and T.S., and NEDO (New Energy and Industrial Technology Development Organization) grants to M.C.S., T.K. and K.A. M.C.S. is supported by CREST from JST. H.S. is a member of the Genome Network Project (MEXT).

Author Contributions Y.K., K.S., T.N.O. and M.C.S. performed AGO2 immunoprecipitations, northern blotting, RNAi, the *in vitro* cleavage assay, β -elimination and qRT-PCR, and prepared the AGO2-associated small RNA library. T.S. characterized and purified the AGO2 antibody. The bioinformatics analyses of AGO2-associated small RNAs were designed and carried out by T.K., K.S., Y.O. and K.A. M.C.S., K.S., Y.K. and H.S. designed the experiments, discussed the interpretation of the results and co-wrote the manuscript.

Author Information Small RNA sequences were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under the accession number GPL6452. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.C.S. (siomim@sc.itc.keio.ac.jp) or H.S. (awa403@sc.itc.keio.ac.jp).

METHODS

Drosophila strains. *Yellow white* (*yw*) was used as a wild-type strain. The strain bearing the *Spn-E*^{100.37} mutation was provided by P. Zamore. The strain bearing the *Dcr-2*^{L811fsX} mutation¹³ was provided by K. Nakahara. The mutant fly (*piggyBac*⁰⁰⁸¹⁰) was used as the *DmHen1/Pimet* mutant. Production of the AGO2 mutant, AGO2⁴¹⁴, was described previously¹².

Cloning of small RNAs associated with AGO2 in S2 cells. AGO1 and AGO2 were immunopurified from S2 cells using specific antibodies against each of the proteins²⁷. Cell lysis buffer was 1× PBS supplemented with 1 mM EDTA, 0.1 mM DTT, 1% Empigen (Calbiochem), 2 µg ml⁻¹ pepstatin, 2 µg ml⁻¹ leupeptin and 0.5% aprotinin. After immunoprecipitation, GammaBind beads (GE Healthcare) that were in a bound form with the immunoprecipitates were washed five times with cell lysis buffer. Total RNAs were isolated from the immunoprecipitates with phenol:chloroform and were precipitated with ethanol. RNAs were dephosphorylated with calf intestinal alkaline phosphatase (CIP) (NEB) and labelled with [γ -³²P]ATP with T4 polynucleotide kinase (Takara) for visualization. Cloning of small RNAs associated with AGO2 in S2 cells was carried out as described^{5,6,9}. The deep sequencing of the libraries was performed on a GS20 system (Roche).

Genome mapping and annotation. Small RNA sequences were mapped to the *D. melanogaster* draft assembly²⁸ (Berkeley *Drosophila* Genome Project release 5) using the NCBI BLASTN program. The annotation of each small RNA was determined by examining the overlap between its mapped regions and other feature track data of the UCSC Genome Browser (<http://genome.ucsc.edu/>). A small RNA was assigned to a feature when the length of its overlap was longer than 90% of the small RNA. We defined the priority of the feature assignment to avoid any conflict of assignment. The assignment to miRNA, rRNA, transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and other non-coding RNAs (ncRNAs) was performed by using the FlyBase²⁹ non-coding track. We used the FlyBase genes track for the annotation of coding genes. We also used our earlier data of piRNAs²² for annotating piRNAs. RepeatMasker (<http://www.repeatmasker.org>) and the natural transposon tracks were used for annotating transposons. Repetitive elements other than transposons were used to assign repeat elements.

Northern blot analysis. Northern blot analysis was carried out essentially as described previously⁵. Total RNAs of S2 cells, fly embryos and adult male flies (carcasses, flies without testes) were isolated using ISOGEN (Invitrogen) according to the manufacturer's instructions. Total RNAs from the immunoprecipitates were isolated with phenol:chloroform and precipitated with ethanol. A DNA fragment for detecting retrotransposon transcripts was cloned into the pBS SK+ vector using primers, 1731-forward and 1731-reverse, for 1731 (GenBank accession number X07656). The sequences of the primers were as follows: 1731-forward, 5'-TATACGGCTTGAAGCAGTCAGGCAG-3'; 1731-reverse, 5'-ACGCCAAGTACCGGAGGATGTGCTT-3'. PCR products were used as templates. DNA probes were synthesized *in vitro* using a random prime labelling kit (Takara) in the presence of [α -³²P]dCTP. Probes used for *miR-279*, *miR-310*, *roo* piRNA, *bantam*, *esiRNA-1731-1* and *esiRNA-sl-1* were as follows; *miR-279*, 5'-ATGAGTGTGGATCTAGTCA-3'; *miR-310*, 5'-AAAGGCCGGG-AAGTGTGCAATA-3'; *roo* piRNA, 5'-TCGACTCAGTGGCACAATAAAT-3'; *bantam*, 5'-CAGCTTTCAAAATGATCTCAC-3'; *esiRNA-1731-1*, 5'-AAGGT-

GTCGTCGCTGGTCTAC-3'; *esiRNA-sl-1*, 5'-GGAGCGAACTTGTTGGA-GTCAA-3'; and U6snRNA, 5'-GGGCCATGCTAATCTTCTCTGTA-3'. The DNA oligonucleotides were labelled with T4 polynucleotide kinase in the presence of [γ -³²P]ATP.

In vitro target-RNA cleavage assay. To make an RNA target harboring sequences completely complementary to *bantam* miRNA and *esiRNA-sl-1*—a short dsDNA fragment (a target sequence for *bantam*) produced from two oligonucleotide DNAs (5'-AGCTTGAGATCATTTTGAAAGCTGATT-3' and 5'-AGCTAATCAGCTTTCAAAATGATCTCA-3') by annealing was first inserted in a pBS SKII+ vector at the HindIII site (yielding pBS-ban). Following this, another DNA fragment (a target sequence for *esiRNA-sl-1*) produced from two DNA oligonucleotides (5'-GATCGGAGCGAACTTGTTGGA-GTCAA-3' and 5'-GGCCTTGACTCCAACAAGTTCGCTCC-3') was inserted into a pBS-ban (between BamHI and NotI). PCR was again performed using primers for the T7 and T3 promoter sequences, and the PCR products used as templates for *in vitro* transcription using a MEGascript T7 kit (Ambion). The resultant RNAs were radiolabelled at their 5' ends with [γ -³²P]ATP using T4 polynucleotide kinase and were gel-purified. The target-RNA cleavage assay was performed as described previously²⁷ using AGO1 and AGO2 immunoprecipitates from S2 in a buffer containing Empigen.

RNA interference. RNAi in S2 cells was performed essentially as described¹². Production of dsRNAs was described previously¹². dsRNAs used in RNAi were: ds*Droscha*, homologous to nucleotides 2701–3925 of the *Droscha* open reading frame (ORF) sequence (accession number AE013599); ds*Dicer-1*, 10–950 of the *Dicer-1* ORF sequence (accession number AE014297); ds*Dicer-2*, 4091–4888 of the *Dicer-2* ORF sequence (accession number NM_079054); dsAGO1, 1035–2042 of the AGO1 ORF sequence (accession number NM_166021); dsAGO2, 1304–2282 of the AGO2 ORF sequence (accession number NM_168626).

Periodate oxidation and β -elimination. Periodate oxidation and β -elimination of RNAs were performed as described^{6,21}. RNAs isolated from AGO1 and AGO2 immunoprecipitates were labelled with [γ -³²P]ATP after CIP treatment. 10,000 counts per min of ³²P-labelled RNAs was used per reaction. Total RNAs of S2 cells and of fly embryos were subjected to periodate oxidation and β -elimination without ³²P labelling. Visualization of *esiRNA-sl-1*, *miR-279*, *roo* piRNA and *miR-310* was done by northern blotting.

Quantitative RT-PCR analysis. Total RNA (0.4 µg) was used to reverse transcribe target sequences using oligo(dT) primer and PrimeScript RT reagent Kit (Takara) according to the manufacturer's instructions. The resulting cDNA was analysed by quantitative RT-PCR in a LightCycler real-time PCR system (Roche Diagnostics) using the SYBR Premix Ex Taq (Takara). Relative steady-state mRNA levels were determined from the threshold cycle for amplification. Supplementary Table 3 lists the PCR primer sequences. Rp49 was used as an internal control.

28. Celniker, S. E. *et al.* Finishing a whole genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Gen. Biol.* **3**, research0079.1–0079.14 (2002).
29. Crosby, M. A., Coodman, J. L., Strelets, V. B., Zhang, P., Gelbart, W. M., the FlyBase Consortium. FlyBase: genomes by the dozen. *Nucleic Acids Res.* **35** (Database issue), D486–D491 (2007).

LETTERS

An endogenous small interfering RNA pathway in *Drosophila*

Benjamin Czech^{1*}, Colin D. Malone^{1*}, Rui Zhou², Alexander Stark^{3,4}, Catherine Schlingeheyde¹, Monica Dus¹, Norbert Perrimon², Manolis Kellis³, James A. Wohlschlegel⁵, Ravi Sachidanandam¹, Gregory J. Hannon¹ & Julius Brennecke¹

Drosophila endogenous small RNAs are categorized according to their mechanisms of biogenesis and the Argonaute protein to which they bind. MicroRNAs are a class of ubiquitously expressed RNAs of ~22 nucleotides in length, which arise from structured precursors through the action of Drosha–Pasha and Dicer-1–Loquacious complexes^{1–7}. These join Argonaute-1 to regulate gene expression^{8,9}. A second endogenous small RNA class, the Piwi-interacting RNAs, bind Piwi proteins and suppress transposons^{10,11}. Piwi-interacting RNAs are restricted to the gonad, and at least a subset of these arises by Piwi-catalysed cleavage of single-stranded RNAs^{12,13}. Here we show that *Drosophila* generates a third small RNA class, endogenous small interfering RNAs, in both gonadal and somatic tissues. Production of these RNAs requires Dicer-2, but a subset depends preferentially on Loquacious^{1,4,5} rather than the canonical Dicer-2 partner, R2D2 (ref. 14). Endogenous small interfering RNAs arise both from convergent transcription units and from structured genomic loci in a tissue-specific fashion. They predominantly join Argonaute-2 and have the capacity, as a class, to target both protein-coding genes and mobile elements. These observations expand the repertoire of small RNAs in *Drosophila*, adding a class that blurs distinctions based on known biogenesis mechanisms and functional roles.

Drosophila melanogaster expresses five Argonaute proteins, which segregate into two classes. The Piwi proteins (Piwi, Aubergine and AGO3) are expressed in gonadal tissues and act with Piwi-interacting RNAs (piRNAs) to suppress mobile genetic elements^{10,11}. The Argonaute class contains AGO1 and AGO2. AGO1 binds microRNAs (miRNAs) and regulates gene expression^{8,9}. The endogenous binding partners of AGO2 have remained enigmatic.

We generated transgenic flies expressing epitope-tagged AGO2 under the control of its endogenous promoter. Tagged AGO2 localized to the cytoplasm of germline and somatic cells of the ovary (Supplementary Fig. 1). Immunoprecipitated AGO2-associated RNAs differed in their mobility from those bound to AGO1 (Fig. 1a). Deep sequencing of small RNAs from AGO1 and AGO2 complexes yielded 2,094,408 AGO1-associated RNAs and 916,834 AGO2-associated RNAs from Schneider (S2) cells, and 455,227 AGO2-associated RNAs from ovaries that matched perfectly to the *Drosophila* genome. We also sequenced three libraries derived from 18–29-nucleotide RNAs (936,833 sequences from wild-type ovaries, 1,042,617 sequences from *Dicer-2* (*Dcr-2*) mutant ovaries, and 1,946,339 sequences from *loquacious* (*loqs*) mutant ovaries) and an 18–24-nucleotide library from wild-type testes (522,848 sequences). Finally, we added to our analysis 92,363

published sequences derived from 19–26-nucleotide RNAs from S2 cells¹⁵.

We noted that among the ~50% of AGO2-associated RNAs from S2 cells that did not match the genome, ~17% matched the flock house virus (FHV), a pathogenic RNA virus and reported target for RNAi in flies^{16,17}. These probably arose because of persistent infection of our S2 cultures.

After excluding presumed degradation products of abundant cellular RNAs, we divided each of the total RNA libraries into two categories: annotated miRNAs and the remainder (Fig. 1b). For the S2 cell library, the size distribution of these populations formed two

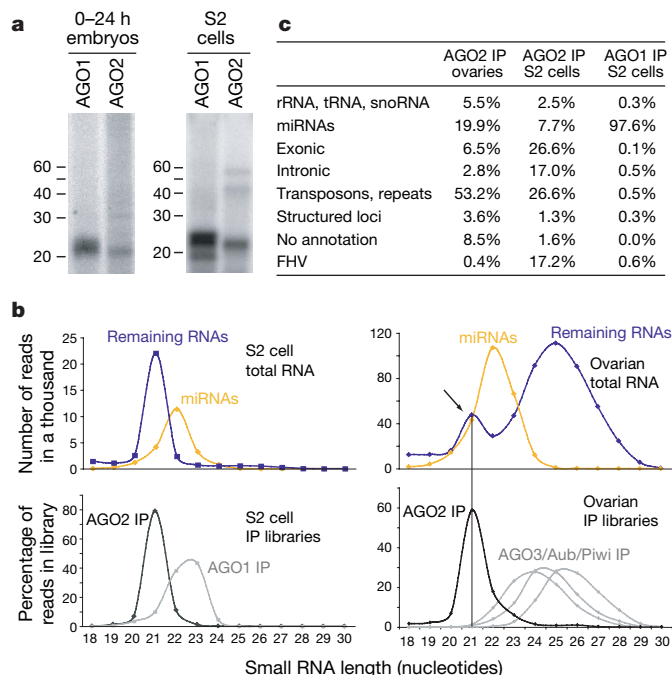


Figure 1 | AGO2 binds endogenous small RNAs. **a**, RNA was isolated from AGO1 and AGO2 immunoprecipitates (IP) from embryos and S2 cells. **b**, Length profiles for small RNAs isolated from S2 cells and ovaries are shown. Species are split into miRNAs and remainder, and are compared to those obtained from individual Argonaute complexes (as indicated). **c**, Annotation of AGO1- and AGO2-associated small RNAs from ovaries and S2 cells. rRNA, ribosomal RNA; snoRNA, small nucleolar RNA; tRNA, transfer RNA.

¹Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²Harvard Medical School, Department of Genetics, Howard Hughes Medical Institute, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ³Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA. ⁴Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Department of Biological Chemistry, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA.

*These authors contributed equally to this work.

peaks, with non-miRNAs lying at 21 nucleotides and miRNAs exhibiting a broader peak from 21 to 23 nucleotides. Libraries derived from AGO1 and AGO2 complexes almost precisely mirrored these two size classes. In the ovary library, this approach revealed three size classes. Whereas two reflected those seen in S2 cells, a third class comprised piRNAs. Again, RNA size profiles from AGO2 or Piwi family immunoprecipitates¹² mirrored those within the total ovary library. These data demonstrate that AGO2 is complexed with a previously uncharacterized population of small RNAs.

Whereas known miRNAs comprised more than 97% of AGO1-associated RNAs in S2 cells, they made up only 8% or 20% of the AGO2-bound species in S2 cells or ovaries, respectively. The remaining small RNAs in AGO2 complexes formed a complex mixture of endogenous siRNAs (endo-siRNAs; Fig. 1c). Among these, transposons and satellite repeats contributed substantially to AGO2-associated small RNAs in S2 cells (27%) and ovaries (53%). The nature of the transposons giving rise to abundant siRNAs in ovaries and S2 cells differed substantially (Fig. 2a), probably reflecting differential expression of specific transposons in these tissues. Unlike piRNAs^{12,13,18,19}, neither somatic nor germline siRNAs exhibited a pronounced enrichment for sense or antisense species (Supplementary Fig. 2a).

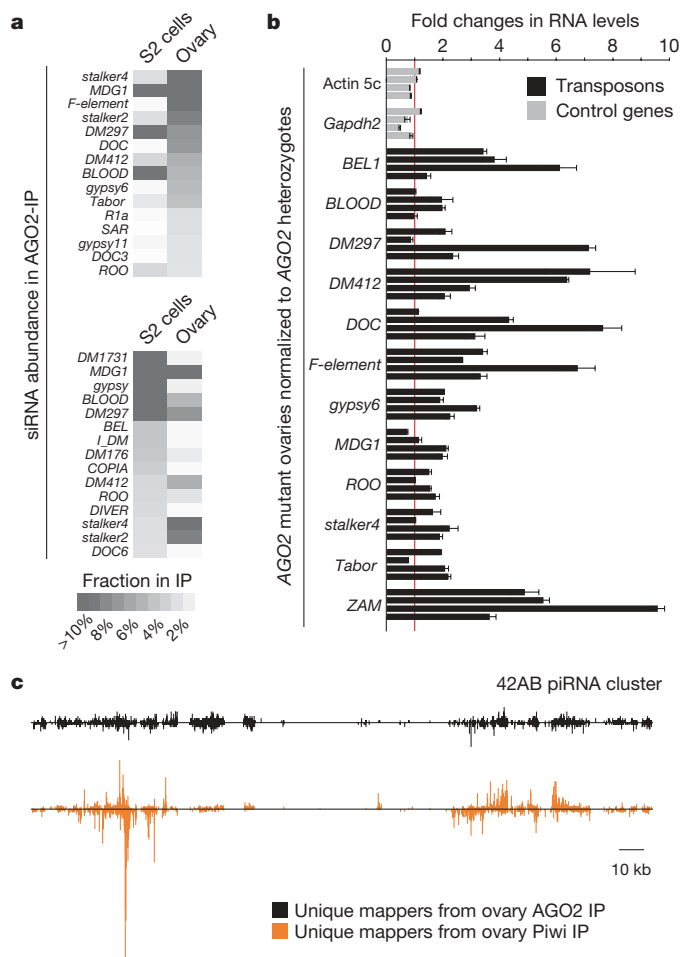


Figure 2 | A subset of endo-siRNAs originates from transposons.

a, Indicated are the cloning frequencies of AGO2-bound siRNAs in ovaries and S2 cells that match individual transposons. **b**, RNA levels of twelve transposons and two control genes in ovaries mutant for AGO2 as compared to AGO2 heterozygotes (four biological replicates; error bars indicate technical variation). **c**, Distributions of AGO2-bound siRNAs (black) and Piwi-bound piRNAs (orange) from ovaries on the piRNA cluster at cytological position 42AB (ref. 12; relative abundances of both populations can be estimated from Supplementary Fig. 2d).

In accord with these findings, knockdown of AGO2 in S2 cells leads to increased expression of several mobile elements²⁰. In the germ line, the Piwi-piRNA system has been reported as the dominant transposon-silencing pathway¹⁹. Nevertheless, we found that several transposons, with a potential to be targeted by siRNAs, were substantially derepressed in AGO2 mutant or *Dcr-2* mutant ovaries (Fig. 2b and Supplementary Fig. 2c). Although comparisons of relative abundance were difficult, both piRNAs and siRNAs mapped to piRNA clusters, with the regions that generate uniquely mapping species generally overlapping (Fig. 2c and Supplementary Fig. 2d). Thus, piRNA loci are a possible source for antisense RNAs matching transposons and might serve a dual function in small RNA generation. Considered together, these data suggest that endo-siRNAs repress the expression of mobile elements, in some tissues acting alongside piRNA pathways.

To probe the nature of the remaining endo-siRNAs, we computationally extracted genomic sites, which give rise to multiple uniquely mapping RNAs that do not fall into heterochromatic regions. These generally segregated into two categories, which we term structured loci and convergently transcribed loci.

Transcripts from structured loci can fold to form extensive double-stranded RNA directly. The two major loci, termed *esi-1* and *esi-2* (Fig. 3a and Supplementary Fig. 3), gave rise to half of the 20 most abundant endo-siRNAs in ovaries and also generated siRNAs in embryos, larvae and adults (not shown). *esi-1*, annotated as CG18854, can produce an ~400-base pair (bp) dsRNA through interaction of its 5' and 3' untranslated regions (UTRs; Supplementary Fig. 3). *esi-2* overlaps with CG4068 and consists of 20 palindromic ~260-nucleotide repeats (Fig. 3a). All siRNAs derived from these two loci arise from one genomic strand. In some previously characterized instances (for example, *Arabidopsis* trans-acting-siRNAs²¹) Dicer generates 'phased' siRNAs with 5' ends showing a 21-nucleotide periodicity. In all tissues examined, *esi-1* and *esi-2* produced phased siRNAs, consistent with a defined initiation site for Dicer processing (Fig. 3a and Supplementary Fig. 3). Phasing was not observed for viral or repeat-derived siRNAs. Finally, siRNAs from both loci also joined AGO1 in proportions greater than siRNAs produced from transposons and repeats, perhaps owing to the imperfect nature of the dsRNA that they produce^{22,23} (Fig. 1c).

AGO2 regulates gene expression by cleavage of complementary sites rather than by recognition of seed sites typical of AGO1-miRNA-mediated regulation²³. We searched for possible targets of endo-siRNAs by identifying transcripts with substantial complementarity. A highly abundant siRNA from *esi-2* is highly complementary to the coding sequence of the DNA-damage-response gene *mutagen-sensitive 308* (*mus308*). Using a modified rapid amplification of cDNA ends (RACE) protocol, we detected *mus308* fragments with 5' ends corresponding precisely to predicted endo-siRNA cleavage sites (Fig. 3b). Moreover, AGO2 and *Dcr-2* loss consistently increased *mus308* expression in testis and to a lesser extent in ovaries, consistent with the relative abundance of *esi-2* siRNAs in these tissues (Fig. 3b, c). Finally, a reporter gene containing two *mus308* target sites was significantly derepressed in S2 cells on depletion of *Dcr-2* or AGO2 but not of *Dcr-1* or AGO1 (Fig. 3c). Although extensive complementarity between other endo-siRNAs and messenger RNAs was rare, we found several *esi-1*-derived siRNAs complementary to CG8289 (Supplementary Fig. 3), suggesting a potential regulatory interaction *in vivo*.

A second group of siRNA-generating loci contained regions in which dsRNAs can arise from convergent transcription. If sorted for siRNA density, most of the top 50 ovarian and S2 cell siRNA loci lay in regions where annotated 3' UTRs or expressed-sequence-tags corresponding to convergently transcribed protein-coding genes overlap (Supplementary Tables 1 and 2). Typically, siRNAs arise on both genomic strands but only from overlapping portions of convergent transcripts (Fig. 3d). Examining all 998 convergently transcribed gene pairs in the *Drosophila* genome with annotated

overlapping transcripts, we found the peak abundance of ovarian siRNAs to be at the centre of the overlap, with sharp declines away from this region (Supplementary Fig. 4). In an alternative arrangement, *Pgant35A* produces sense and antisense siRNAs across its entire annotated transcript, consistent with expressed-sequence-tag support for antisense transcription traversing this locus (Supplementary Fig. 5).

Thus, a large number of *Drosophila* genes generate endogenous siRNAs, with most having perfect complementarity to the 3' UTRs of neighbouring genes. Relative levels of endo-siRNAs generated from each convergent transcription unit were low (not shown), and we found no or little change (up to a ~1.3-fold increase) in the expression of such genes in AGO2 mutant ovaries. Possibly, the level of small RNAs produced by this genomic arrangement is

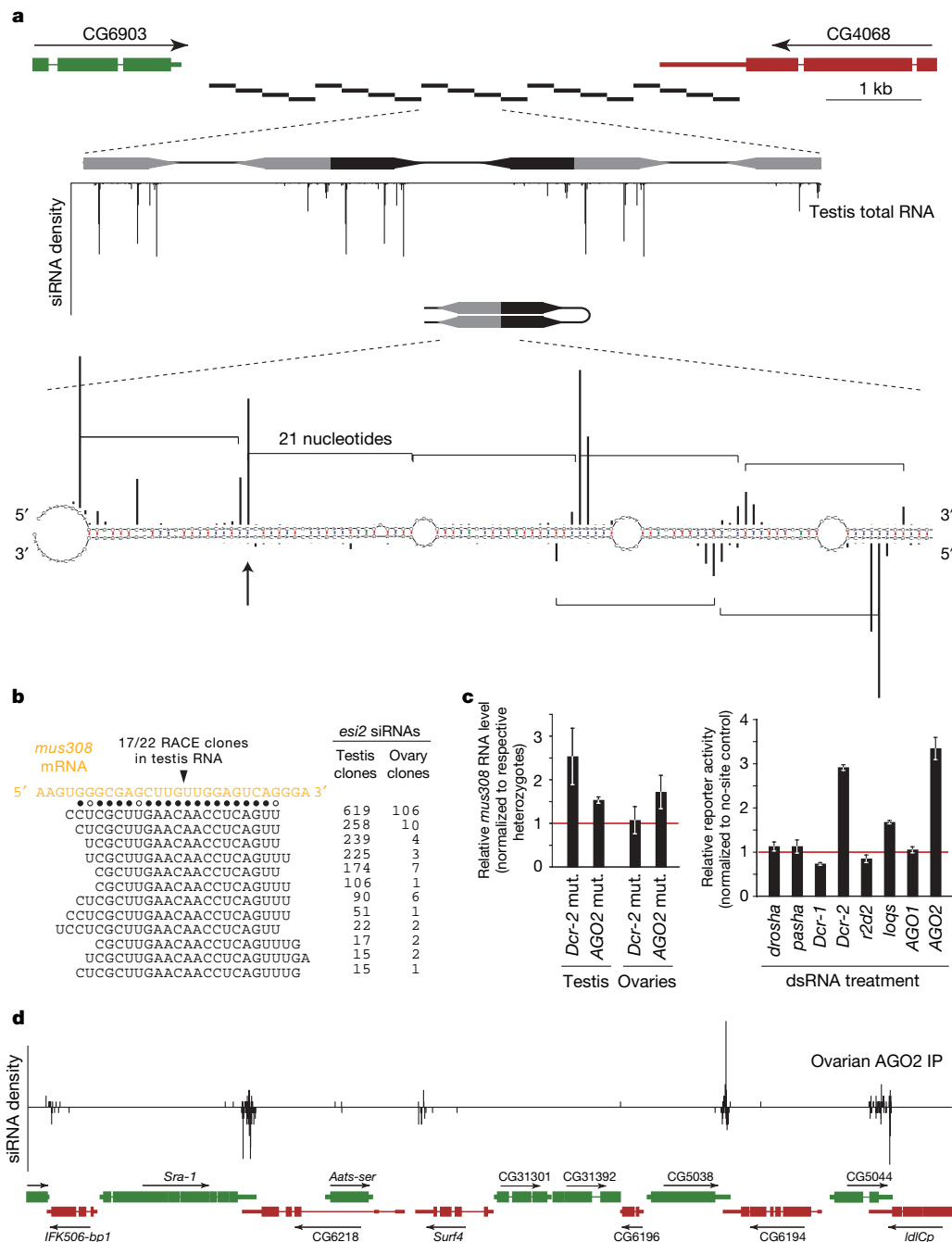


Figure 3 | Two types of genic endo-siRNA loci in *Drosophila*. **a**, FlyBase gene structure at the *esi-2* locus indicating the twenty ~260-bp repeats of *esi-2* (black bars). Below, three repeats are magnified and the small RNA density is depicted. Most siRNAs match up to 20 times and cannot be unambiguously assigned to a chromosomal location. Each repeat is a palindromic sequence, offering a multitude of possible structures. One structure is shown with the 5' ends of cloned siRNAs as black bars (the height correlates with cloning frequency; phasing is indicated by brackets). **b**, An abundant siRNA from *esi-2* (indicated by an arrow in **a**) is highly complementary to the *mus308* mRNA. Shown is the *mus308* target site and cloned siRNAs (solid dots indicate canonical base pairs; open dots indicate

GU base pairs). The only detected cleavage site within this duplex is indicated above. **c**, Shown are *mus308* transcript levels from AGO2 and *Dcr-2* mutant (mut.) flies compared to their respective heterozygotes (error bars indicate standard deviation). To the right, average reporter levels (error bars indicate standard deviation) of a construct containing two *mus308* target sites in S2 cells depleted of the indicated genes by RNAi are shown. **d**, An example of siRNAs arising from convergent transcription units. A 30-kb region containing multiple instances of convergently transcribed genes is displayed, with the density of AGO2-associated RNAs in ovaries shown above.

inconsequential, amounting to noise within silencing pathways. However, there are probably circumstances wherein regulation by such arrangements might substantially impact expression.

In S2 cells, two neighbouring loci encoded nearly 16% of AGO2-associated RNAs (Supplementary Table 2). These reside within a large intron of *klarsicht* (Supplementary Fig. 6) and did not generate siRNAs in any other tissue. A similar locus, corresponding to CG14033, was found within an intron of *thickveins* (Supplementary Fig. 7) and gave rise to testis-specific siRNAs. Although the function of both siRNA clusters is unclear, the *thickveins* cluster shares considerable complementarity to CG9203, and loss of AGO2 and *Dcr-2* mildly increased CG9203 mRNA levels in testis but not in ovaries (Supplementary Fig. 7).

Dcr-2 has been implicated in the production of siRNAs from viral replication intermediates or exogenously introduced dsRNAs, whereas *Dcr-1* has been linked to miRNA biogenesis^{6,16,17}. In agreement with these observations, all endo-siRNA classes were lost in *Dcr-2* mutant ovaries (Fig. 4a). To obtain more insight into the genetic requirements for endo-siRNA biogenesis and stability, we depleted components of siRNA and miRNA pathways in S2 cells and analysed levels of abundant siRNAs derived from structured loci (Fig. 4b and Supplementary Fig. 8). Although depletion of *Dcr-2* and AGO2 resulted in substantial reductions in siRNA levels, little or no changes were observed on Droscha, Pasha, *Dcr-1* or AGO1 depletion. Unexpectedly, we found virtually no requirement for the *Dcr-2* partner R2D2 (ref. 14) but a strong requirement for the *Dcr-1* partner Loquacious^{1,4,5}. Only one analysed siRNA exhibited partial

dependence on R2D2, potentially correlating with the extensive dsRNA character of its precursor duplex (Supplementary Fig. 9). Artificial sensors for endo-siRNAs from *esi-1* and *esi-2* in S2 cells gave patterns of de-repression that matched our analysis of endo-siRNA levels (Figs 3c and 4c).

Analysis of the most abundant siRNA from *esi-2* in flies mutant for *Dcr-2*, AGO2, *r2d2* or *loqs* extended our findings from cell culture (Supplementary Fig. 10). To examine the unexpected requirement for *loqs* more broadly, we sequenced small RNAs from *loqs*-mutant ovaries and observed a near complete loss of endo-siRNAs from structured loci (Fig. 4a). A much smaller impact of *loqs* was seen on endo-siRNAs derived from repeats and convergent transcription units. However, an involvement of Loqs and not R2D2 in the function of siRNAs derived from perfect dsRNA precursors was supported by analysing the impact of depleting siRNA/miRNA pathway components on the ability to suppress FHV replication in our infected S2 cell cultures (Supplementary Fig. 11).

Our results uncover an unanticipated role for Loqs in siRNA biogenesis and suggest that R2D2 has a lesser impact on at least two types of endogenous siRNAs. It is well established that Loqs partners with *Dcr-1* for miRNA processing. To probe a molecular interaction with *Dcr-2*, we catalogued Loqs binding partners using quantitative proteomics. *Dcr-1* and *Dcr-2* were both abundant in Loqs immunoprecipitates from cultured cells and flies (Supplementary Fig. 12), supporting a physical interaction between *Dcr-2* and Loqs.

Among animals, endo-siRNA pathways have so far been restricted to *Caenorhabditis elegans*^{24–27}. Our results extend the prevalence of

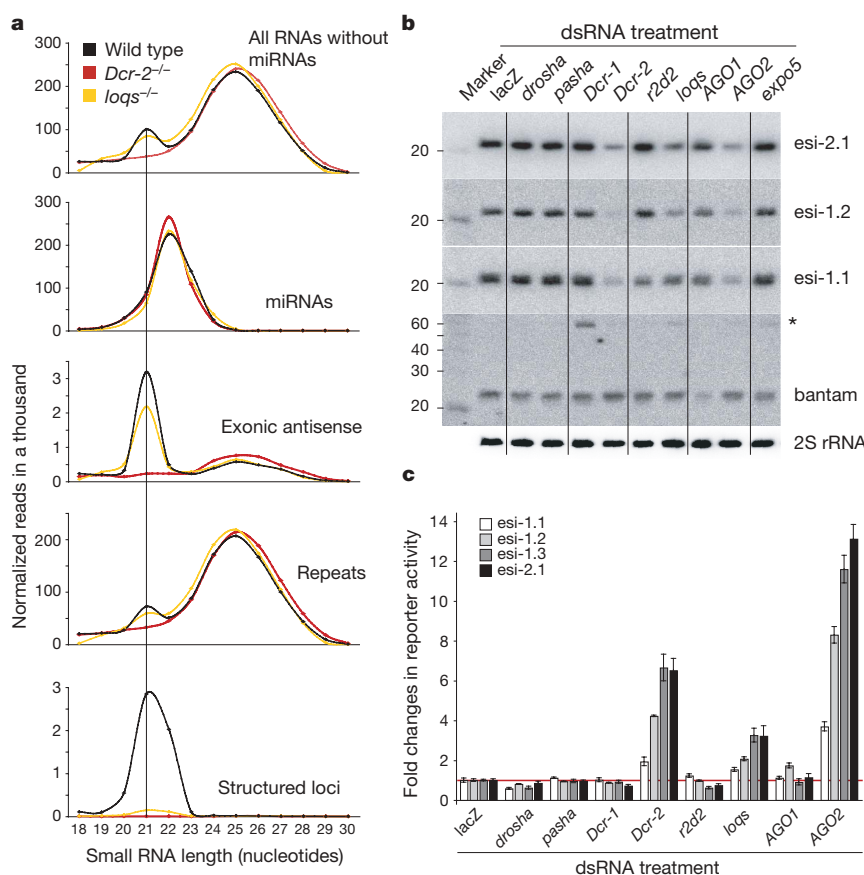


Figure 4 | Genetic requirements for siRNA biogenesis. **a**, Length distributions of small RNAs from total RNA libraries obtained from wild-type (black), *Dcr-2* mutant (red) and *loqs* mutant (yellow) ovaries. Excluding miRNAs and piRNAs (23–29 nucleotides), the population of 21-nucleotide siRNAs mapping to structured loci, genes and repeats is lost from *Dcr-2* mutants and those mapping to structured loci are strongly reduced in *loqs* mutants (all libraries are adjusted to the same total small RNA count).

b, Northern blots showing levels of three siRNAs encoded from structured loci *esi-1* and *esi-2* (esi-2.1, esi-1.1 and esi-1.2) in S2 cells treated with dsRNA against the genes indicated. As controls, northern blots of *bantam* (pre-miRNA indicated by the asterisk) and 2S rRNA are shown below. **c**, *Renilla* luciferase reporter assays are shown for the siRNAs examined in **b** and an additional esi-1-derived species (esi-1.3) in S2 cells treated with dsRNAs against the indicated genes (error bars indicate standard deviation; $n = 3$).

such systems to *Drosophila* and parallel recent discoveries of an endo-siRNA pathway in mouse oocytes^{28,29}. These systems have many common features but also key differences. In both, siRNAs collaborate with piRNAs to repress transposons. Also, mouse and *Drosophila* both generate endo-siRNAs from structured loci. In mouse, dsRNAs can form by pairing of sense protein-coding transcripts with antisense transcripts from pseudogenes. Whether or not transcripts from unlinked sites lead to siRNA production in *Drosophila* is unclear. However, transposon sense transcripts may hybridize to antisense sequences transcribed from piRNA clusters to form endo-siRNA precursors. In flies, a much larger number of genic loci enter the pathway as compared to mice because convergent transcription of neighbouring genes frequently creates overlapping transcripts. Overall, annotation of the *Drosophila* genome indicates that a significant proportion is transcribed in both orientations, providing widespread potential for dsRNA formation. This property is shared by many other annotated genomes, raising the possibility that the RNAi pathway has broad impacts on gene regulation. Viewed in combination, our studies suggest an evolutionarily widespread adoption of dsRNAs as regulatory molecules, a property previously ascribed only to miRNAs.

METHODS SUMMARY

The fly stocks used were *Dcr-2^{L811Fxx}* (ref. 6), *AGO2⁴¹⁴* (ref. 30), *log^{s00791}* (ref. 1) and *r2d2¹* (ref. 14). Recombineering was used to insert a Flag-haemagglutinin (HA) tag at the amino terminus of the AGO2 coding sequence in the context of the genomic AGO2 locus including flanking regulatory regions (for details, see Methods). Polyclonal anti-AGO1 antibody was obtained from Abcam (lot number 113754). Small RNAs for library production were isolated from ovarian total RNA or from Argonaute immunoprecipitates. Libraries were produced as described¹² and sequenced using the Illumina platform (protocol available on request). A description of the bioinformatics methods can be found online. For quantitative real-time PCR (qRT-PCR) analyses, we used total RNA preparations and random hexamer primers. Details and all primer sequences are given in the Supplementary Information. S2 cell knockdown treatments were for eight days with two sequential dsRNA soakings. For the reporter experiments, inducible expression plasmids for *Renilla* and firefly were transfected into S2 cells together with dsRNA for the desired knockdown target. *Renilla* constructs contained two target sites for endogenous siRNAs, whereas the firefly construct was used for normalization. For details on plasmids, dsRNAs and target sites, see Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 January; accepted 18 April 2008.

Published online 7 May 2008.

- Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
- Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
- Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).
- Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
- Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
- Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).

- Eulalio, A., Huntzinger, E. & Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. *Cell* **132**, 9–14 (2008).
- Bushati, N. & Cohen, S. M. microRNA functions. *Annu. Rev. Cell Dev. Biol.* **23**, 175–205 (2007).
- Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
- Klattenhoff, C. & Theurkauf, W. Biogenesis and germline functions of piRNAs. *Development* **135**, 3–9 (2008).
- Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
- Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
- Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
- Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17**, 1850–1864 (2007).
- Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A. & Imler, J. L. Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nature Immunol.* **7**, 590–597 (2006).
- Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452–454 (2006).
- Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20**, 2214–2222 (2006).
- Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
- Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **26**, 2965–2975 (2006).
- Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
- Tomari, Y., Du, T. & Zamore, P. D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**, 299–308 (2007).
- Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
- Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
- Sijen, T., Steiner, F. A., Thijssen, K. L. & Plasterk, R. H. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**, 244–247 (2007).
- Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241–244 (2007).
- Yigit, E. *et al.* Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**, 747–757 (2006).
- Tam, O. H. *et al.* Pseudogene-derived siRNAs regulate gene expression in mouse oocytes. *Nature advance online publication*, doi:10.1038/nature06904 (10 April 2008).
- Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature advance online publication*, doi:10.1038/nature06908 (10 April 2008).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Carthew, H. Siomi, P. Zamore and D. Smith for reagents. We are grateful to M. Rooks, E. Hodges and D. McCombie for help with deep sequencing. B.C. was supported by the German Academic Exchange Service. C.D.M. is a Beckman fellow of the Watson School of Biological Sciences and is supported by an NSF Graduate Research Fellowship. R.Z. is a Special Fellow of the Leukemia and Lymphoma Society. M.D. is an Engelhorn fellow of the Watson School of Biological Sciences. J.B. is supported by the Ernst Schering foundation. A.S. is supported by an HFSP fellowship. This work was supported in part from grants from the NIH to G.J.H. and N.P. and a gift from K. W. Davis (G.J.H.).

Author Information Small RNA sequences were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE11086. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu) or J.B. (brenneck@cshl.edu).

METHODS

Fly stocks. A 3× Flag–HA tag was inserted at the N terminus of the Flybase RB transcript of AGO2 into BAC RP98-21A13 by means of bacterial red/ET recombination (Gene Bridges GmbH). A 13.9-kb AvrII/XhoI fragment of the modified BAC encompassing the AGO2 locus including parts of flanking genes (chromosome 3L coordinates: 15,544,405–15,558,309) was cloned into pCasper4 (XbaI/XhoI). Transgenic flies were generated at Bestgene Inc. Expression of tagged AGO2 in embryos, ovaries and whole flies was verified in multiple lines by western blotting using a monoclonal anti-HA-peroxidase antibody (1:500; catalogue number 12013819001, Roche). *Dcr-2^{L811F^{sc}}* flies were a gift of R. Carthew⁶, *AGO2⁴¹⁴* flies were a gift of H. Siomi³⁰, *loqs¹⁰⁰⁷⁹¹* flies were a gift of P. Zamore¹ and *r2d2¹* flies were a gift of D. Smith¹⁴. For wild-type fly stocks, stock number 2057 from Bloomington (Celera sequencing strain) and OregonR flies were used.

Small RNA libraries. Twenty-six 10-cm plates with 50–70% confluent Schneider cells were transfected with pCasper_Flag-HA-AGO2 using calcium phosphate, harvested after 36 h and lysed in buffer A (20 mM HEPES, pH 7.0, 150 mM NaCl, 2.5 mM MgCl₂, 0.3% Triton, 30% glycerol) supplemented with 1 mM PMSF and protease inhibitors (Complete, Roche). Cleared extract was split and incubated with rabbit polyclonal AGO1 antibody (1:20; lot number 113754, Abcam) or mouse anti-Flag M2-agarose (1:25, Sigma) for 4 h at 4 °C. AGO1 antibodies were isolated by adding protein G beads (1:10; Roche) for 1 h. Beads were washed six times each for 10 min in buffer B (30 mM HEPES, pH 7.4, 800 mM NaCl, 2 mM MgCl₂, 0.1% NP-40), which contained equal supplements as buffer A. The immunoprecipitation was analysed by western blotting using anti-AGO1 (1:2,000; Abcam) and anti-HA-peroxidase (1:500; catalogue number 12013819001, Roche). AGO1- and AGO2-associated RNAs were isolated with phenol/chloroform and ethanol precipitated. For the ovarian AGO2 IP library, ~500 mg ovaries from transgenic Flag-HA-AGO2 flies were dissected and lysed mechanically in buffer A. AGO2 complexes and associated RNAs were purified as above.

AGO1- and AGO2-bound small RNAs as well as small RNAs from total RNA were cloned as described¹² (the detailed protocol is available on request). The following small RNA libraries from total RNA were prepared for this study: 18–28 nucleotides from ovaries of the Celera sequenced strain (Bloomington, number 2057); 18–28 nucleotides from ovaries of *Dcr-2^{L811F^{sc}}* homozygous flies; 18–28 nucleotides from *loqs¹⁰⁰⁷⁹¹* homozygous flies; and 18–24 nucleotides from the testis of OregonR flies.

Libraries were sequenced in house using the Illumina platform. Published libraries used in this study were a 16–26-nucleotide S2 cell total RNA library¹⁵ and libraries from ovarian Piwi/Aub/AGO3 immunoprecipitates¹².

Bioinformatic analysis of small RNA libraries. Small RNA sequences were matched to the *Drosophila* release 5 genome and to genomes of *Drosophila C* virus, FHV and cricket paralysis virus. Only reads matching the fly genome 100% and viral genomes with up to three mismatches were used for further analysis. For annotations, we used Flybase for protein-coding genes, UCSC for non-coding RNAs and transposons/repeats (<http://genome.ucsc.edu/>) and the most recent miRNA catalogue^{15,31}.

siRNA clusters were extracted by mapping all 20–22-nucleotide-long RNAs from the AGO2-IP libraries to the genome (only uniquely mapping RNAs were used) and retaining 200-nucleotide windows that contained at least three distinct small RNAs. Windows separated by a maximum of 200 nucleotides were fused and those with more than 40 unique reads were sorted after the density of siRNAs per base pair.

For the transposon analysis, 20–22-nucleotide AGO2-bound RNAs from ovaries and S2 cells were mapped onto the Repbase collection of transposons³² with up to three mismatches to construct heatmaps indicating cloning frequency and strand bias of siRNAs. For the latter analysis, only siRNAs unambiguously mapped to one strand were considered.

Cleavage site mapping for endo-siRNA targets. Wild-type testes were dissected on ice into 1× PBS. Total RNA was isolated using Trizol (Invitrogen) according to the manufacturer's protocol. Total RNA (5 µg) was used as starting material. Ligation of an RNA adaptor, reverse transcription using the GeneRacer oligo(dT) primer and 5' RACE-PCR were performed according to the manufacturer's instructions (GeneRacer kit, Invitrogen). 5' RACE-PCR was carried out using the GeneRacer 5' primer (5'-CGACTGGAGCAGGAGCACTGA-3') and a *mus308* gene-specific reverse primer (5'-TGCTTTGCAGAGTCGAA-GCTGATTG-3'), and followed by one round of nested PCR using the GeneRacer 5' nested primer (5'-GGACACTGACATGGACTGAAGGAGTA-3') and a nested primer specific to *mus308* (5'-CCGCTAGCTCTACCAAACTGGTGAT-3'). PCR products were gel purified and cloned into pCR4Blunt-TOPO (Invitrogen). Twenty-two clones were sequenced with T7 (5'-GTAATACGACTCACTATAGGGC-3') and T3 (5'-AATTAACCCTCACTAAAGGG-3') primers, and subjected to further analysis.

dsRNA treatment of Schneider cells. Approximately 3×10^6 S2-NP cells were soaked in 1.5 ml serum-free Schneider's medium containing 10 µg dsRNAs in 6-well plates, and 3 ml serum-containing medium was added 45 min later. After 4 days of initial dsRNA treatment, cells were treated with a second round of dsRNAs using the same procedure, and were harvested another 4 days later. Total RNA was extracted with Trizol (Invitrogen). Sequences of the primers for generating dsRNAs are listed below.

siRNA reporter constructs. A SalI/BglII fragment from pGL3-Basic (Promega) was cloned to pRmHa-3 using SalI/BamHI (pMT-Firefly-long). The coding region of the *Renilla* luciferase gene was amplified by PCR and cloned into pRmHa-3 using BamHI/EcoRI sites (pMT-Renilla). A pair of oligonucleotides containing two perfect binding sites for si1_1, si1_2 or si2 were annealed and cloned into pMT-Renilla (BamHI/SalI) to generate sensor constructs (Supplementary Fig. 13).

Transfection was performed in a 384-well plate format. For each well, ~100 ng plasmid DNA (5 ng pMT-Renilla, 20 ng pMT-Renilla-si1_1, 50 ng pMT-si1_2 or 100 ng pMT-si2, 5 ng pMT-Firefly-long, and corresponding amounts of pRmHa-3 serving as carrier DNA) and ~80 ng dsRNA were mixed with 0.8 µl enhancer in 15 µl EC (Qiagen) and the mixture was incubated at room temperature (23 °C) for 5 min. After this, 0.35 µl of Effectene reagent was added and the mixture was immediately dispensed into each well containing dsRNA. After incubation at room temperature for 10 min, 40 µl S2-NP cells (10^6 cells ml⁻¹) were dispensed into the well. Cells were induced with 200 µM CuSO₄ 132 h post transfection, and luciferase assays were performed 36 h later using DualGlo reagents (Promega). For each well, the reporter activity was calculated as the ratio of *Renilla* luciferase to firefly luciferase. Each data point was normalized against the data points where dsRNA against LacZ was transfected. Presented are average results with standard deviation ($n = 3$).

Northern blotting. Total RNA was isolated using Trizol (Invitrogen). RNA (30 µg) was separated on a 15% denaturing polyacrylamide gel and transferred onto a Hybond-N+ membrane (Amersham Biosciences) in 1× TBE (Tris-Borate-EDTA) buffer. The RNA was crosslinked using ultraviolet light (Stratalinker) to the membrane and pre-hybridized in ULTRAhyb buffer (Ambion) for 1 h. DNA probes complementary to the indicated endo-siRNAs, bantam and 2S rRNA were 5' radiolabelled and added to the hybridization buffer (hybridization overnight at 37 °C). Membranes were washed 4–6 times in 1× SSC with 0.1% SDS at 37 °C and exposed to PhosphorImager screens. Probes were stripped by boiling the membrane twice in 0.2× SSC containing 0.1% SDS in a microwave.

Quantitative real-time PCR. Ovaries and testis from homozygous or heterozygous flies were dissected on ice into 1× PBS. Total RNA of dissected tissues or S2 cells was extracted using Trizol (Invitrogen). RNA was treated with DNase I Amplification Grade (Invitrogen) according to the manufacturer's instructions. Complementary DNA was prepared by reverse transcription using SuperScript III Reverse Transcriptase (Invitrogen) and random hexamer primer. qRT-PCR was carried out using SYBR GREEN PCR Master Mix (Applied Biosystems) and a Chromo4 Real-Time PCR Detector (BioRad). C_t values were calculated within the log-linear phase of the amplification curve using the Opticon Monitor 3.1.32 software (BioRad). Quantification was normalized to the mRNA coding for the endogenous ribosomal protein rp49, and relative expression levels were calculated using the following equation: $A = 1.8^{[C_t(\text{ref}) - C_t(\text{ref-control})] - [C_t(\text{sample}) - C_t(\text{sample-control})]}$. Transposon analysis was carried out with four biological replicates (individually shown; error bars indicate technical replicates); siRNA target analysis was carried out with three biological replicates. Oligonucleotide primers used in this study are listed below.

DNA oligonucleotides. PCR primers for the generation of dsRNAs were as follows: T7-Dicer-1-F-14 TAATACGACTCACTATAGGGTGCACAACAA-TCTGC; T7-Dicer-1-R-565 TAATACGACTCACTATAGGGTCACTTGCTG-CAGCTCAC; T7-Dicer-2-F-5 TAATACGACTCACTATAGGGAAGATGTGG-AAATCAAGCC; T7-Dicer-2-R-555 TAATACGACTCACTATAGGGCCACG-TTCGTAATTTTC; T7-Drosha-F-3356 TAATACGACTCACTATAGGGTGAA-TCAGGACTGGAACG; T7-Drosha-R-3910 TAATACGACTCACTATAGGG-AGCCATCGCTATCACTGC; T7-Exportin5-F-55 TAATACGACTCACTATA-GGGATCTAGTCATGAACCCG; T7-Exportin5-R-623 TAATACGACTCACT-ATAGGGAACGAGTCACATGCTGC; T7-AGO1-F1225 TAATACGACTC-ACTATAGGGAACGACAGACCGTAGAG; T7-AGO1-R1858 TAATACGAC-TCACTATAGGGTGGCGTACTTACAGAAGC; T7-AGO2-F2211 TAATACG-ACTCACTATAGGGAGCCACATCGACGAACG; T7-AGO2-R2855 TAATAC-GACTCACTATAGGGGAGGAGTATCCTTGATC; T7-R2D2-PZ-F TAAT-ACGACTCACTATAGGGCATAACCGCTTGATGAAGGATTC; T7-R2D2-PZ-R TAATACGACTCACTATAGGGTGTGCTGCTCGCTACTTGC; T7-Pasha-F452 TAATACGACTCACTATAGGGACTTTGAAGTCTACCCG; T7-Pasha-R1177 TAATACGACTCACTATAGGGGTCCTTGAAGTCACTAGG; T7-Loqs-F-1 TAATACGACTCACTATAGGGATGGACCAGGAGAATTTC; T7-

Loqs-R-540 TAATACGACTCACTATAGGGAAGGGCGTATCCTTGTC; T7-LacZ-F TAATACGACTCACTATAGGGCATTATCCGAACCATCC; T7-LacZ-R TAATACGACTCACTATAGGGCAGAACTGGCGATCGTTCCG. endo-siRNA sensor oligonucleotides were as follows: BamHI-esi-1_2-S2-2P-F, GATCCCAACAGTTTATTTACTTGGAGGCAACATAATCAAATGAACTGAGGGTTACTTGGAGGCAACATAATCAG; Sall-esi-1_2-S2-2P-R, TCGACTGATTATGTTGCCCTCCAAGTAAACATAAAGTGTGG; BamHI-esi-2_1-S2-2P-F, GATCCCAACAGTTTATTGGAGCGAACTTGTGGAGTCAAG; Sall-esi-2_1-S2-2P-R, TCGACTTGACTCCAACAAGTTCGCTCCACCCTCAGTTCATTTGACTCCAACAAGTTCGCTCCAATAAACTGTTGG; BamHI-esi-1_3-S2-2P-F, GATCCCAACAGTTTATTCATTTGATCCATAGTTTCCCGAATGAACTGAGGGTCATTTGATCCATAGTTTCCCGG; Sall-esi-1_3-S2-2P-R, TCGACCGGAAACTATGGATCAAATGACCCTCAGTTCATTCCGGAAACTATGGATCAAATGAATAAACTGTTGG; BamHI-esi-1_1-S2-2P-F, GATCCCAACAGTTTATTGCCAAGGTACGTGGTCGACCGAAATGAACTGAGGGTGCCAAGGTACGTGGTCGACCGAG; Sall-esi-1_1-BC36-S2-2P-R, TCGACTCGGTGACACGTACCTTGGCACCCTCAGTTCATTTGTCGACCACTGACCTTGGCAATAAACTGTTGG; BamHI-mus308_target-S2-2P-F, GATCCCAACAGTTTATTGGGCGAGCTTGTGGAGTCAGAATGAACTGAGGGTGGGCGAGCTTGTGGAGTCAGG; Sall-mus308_target-S2-2P-R, TCGACCTGACTCCAACAAGCTCGCCACCCTCAGTTCATTCTGACTCCAACAAGCTCGCCCAATAAACTGTTGG.

Northern probes were as follows: esi-2.1, GGAGCGAACTTGTGGAGTCAA; esi-1.1, GCCAAGGTACGTGGTCGACCGA; esi-1.2, CATTTGATCCATAGTTTCCCG; miR-bantam, AATCAGCTTCAAAATGATCTCA; 2S rRNA, TACAACCCTCAACCATATGTAGTCCAAGCA.

Quantitative real-time PCR. For transposon analysis: rp49_F, ATGACCATCCGCCCCAGCATAC; rp49_R, CTGCATGAGCAGGACCTCCAG; GAPDH2_F, TGATGAAATTAAGGCCAAGGTTTCAGGA; GAPDH2_R, TCGTTGTCTGTACCAAGAGATCAGCTTC; actin5c_F, AAGTTGCTGCTCTGGTTGTCTG; actin5c_R, GCCACACGCAGCTCATTGTAG; BEL1_F, ATTATACAAACGCCCAATTGCCAAAA; BEL1_R, TCCGATGAAGCTGCAGACAAATAAGA; BLOOD_F, AGACGTTTACATACAGATCAAGGTACGGA; BLOOD_R, AGTTCGTATGGCAATAGTCATGGACT; DM412_F, AAAGTACGGTCCAATGAGAGC; DM412_R, GTGGTGATGAGCTGTTGATGTT; F-element_F, TTGTTGAACAGCATACCACTCC; F-element_R, CCAGAGTTGATGAGCCAGTGTA; gypsy6_F, GACAAGGGCATAACCGATACTGTGGA; gypsy6_R, AATGATTCTGTTCGGGACTTCCGTCT; MDG1_F, AACAGAAACGCCAGCAACAGC; MDG1_R, CGTTCCTCATGTCCGTTGTGAT; ROO_F, CGTCTGCAATGTACTGGCTCT; ROO_R, CGGCACTCCACTAAGTCTCTCC; stalker4_F, TTTGGAAGATTACCAAGGCAGTTCGC; stalker4_R, GGATCTAAGTATGACCCGATTCTGTC; ZAM_F, ACTTGACCTGGATACACTCACAAC; ZAM_R, GAGTATTACGCGGACTAGGGATAC; FHV_F, CCCTGGAGTCGCTTACTTGAGTGCT; FHV_R, ATGGAAGCGTACCTGAAGGAGGACA; DOC_F, TACCTTAAACAAACAACATGCCACC; DOC_R, TTTGTATGGGTGGTCAGCTTTTCGT; DM297_F, GCCAGTACACAGCAACGAAATA; DM297_R, AATTGAATTTTGCAATTTTGG; Tabor_F, GAGCAAGAATTATGCTCGAAGAA; Tabor_R, AATTTATGTCCGGTTTCGTTTTT.

For endo-siRNA target analysis: mus308_F, AAGGATTAGCGCCAAGCTGGAGGAT; mus308_R, ACCACGACCACTGCCACAGAGATT; CG9203_F, AGCTGGCAGAAAAACCATGACCAGT; CG9203_R, CAATTCCTTTGGCGTAGCTTGAGCA.

For S2 cell knockdown analysis: S2-Dcr-1_F, ACGCCTTCCATCTCCCAGTTTACC; S2-Dcr-1_R, GCCACCCTGCTTATTCTGACTGCTC; S2-Dcr-2_F, AAACGAGAGATTCGTGCCCAAAACA; S2-Dcr-2_R, CTGTCCTTGCTCTTATCGGCCTTGT; S2-Drosha_F, AGATGCCAGAGAACTTCACCATCCA; S2-Drosha_R, GAAAGAAGTGAAGAGCTGGGCAGGA; S2-Pasha_F, TGTCAAGGACAAAGATAACGGGCAACA; S2-Pasha_R, GTTGGGAGATGGCTCCGTCT; S2-AGO1_F, ACTACCACGTTCTGTGGGACGACAA; S2-AGO1_R, GAATCGTGCTCTTCTCCACCAGAT; S2-AGO2_F, AACCTCAAAAGTAATCATGGGAAA; S2-AGO2_R, ATTTTGTGCTGTTGGCCTCCTTG; S2-Loqs_F, GTGTGTGCTCTGGATTTGCTGTA; S2-Loqs_R, GTTTTCGGGAGGATTCGGTGTGAT; S2-R2D2_F, GCGAAGACGGAGGGTACGTCTGTAA; S2-R2D2_R, AGTCGAATCCTTCATCAAGCCGTGT.

31. Stark, A. *et al.* Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* 17, 1865–1879 (2007).
32. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467 (2005).

The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs

Katsutomo Okamura¹, Wei-Jen Chung¹, J. Graham Ruby², Huili Guo², David P. Bartel² & Eric C. Lai¹

In contrast to microRNAs and Piwi-associated RNAs, short interfering RNAs (siRNAs) are seemingly dispensable for host-directed gene regulation in *Drosophila*. This notion is based on the fact that mutants lacking the core siRNA-generating enzyme Dicer-2 or the predominant siRNA effector Argonaute 2 are viable, fertile and of relatively normal morphology^{1,2}. Moreover, endogenous *Drosophila* siRNAs have not yet been identified. Here we report that siRNAs derived from long hairpin RNA genes (hpRNAs) programme Slicer complexes that can repress endogenous target transcripts. The *Drosophila* hpRNA pathway is a hybrid mechanism that combines canonical RNA interference factors (Dicer-2, Hen1 (known as CG12367) and Argonaute 2) with a canonical microRNA factor (Loquacious) to generate ~21-nucleotide siRNAs. These novel regulatory RNAs reveal unexpected complexity in the sorting of small RNAs, and open a window onto the biological usage of endogenous RNA interference in *Drosophila*.

Artificial, long-inverted repeat transcripts are efficiently processed by a Dicer-2 (Dcr-2)/Argonaute 2 (AGO2)-driven RNA interference (RNAi) pathway in transgenic *Drosophila*^{1,3}. We hypothesized that this might reflect the existence of an endogenous pathway that accepts long, inverted repeat transcripts. To test this idea, we searched for inverted repeats using EINVERTED⁴ and selected putative hairpins containing mapped small RNA reads (see Methods). Out of 8,132 candidate regions, most consisted of the terminal inverted repeats of individual transposable elements or long terminal repeats of tandem inverted transposable elements. The remaining loci corresponded to inverted tandem duplications of messenger RNA- or transfer RNA-encoding genes, a microRNA (miRNA) gene (*mir-997*), a novel tandem pair of short hairpins (*chou39-1* and *chou39-2*, Supplementary Fig. 1), and a handful of single-gene annotations and unannotated regions.

We analysed the size distribution of cloned RNAs from all of the non-transposable-element EINVERTED hits. Although these mostly exhibited a broad length distribution across the ~18–26-nucleotide cloning range, indicative of degradation fragments (Supplementary Fig. 1b), seven genomic regions specifically generated 21–22-nucleotide RNAs (Supplementary Fig. 1c–f). These included genes annotated as CG18854, CG32207, CR32205 and pncr009 (also known as pncr009:3L), a series of 20 repeats that partially overlap the 3' untranslated region (UTR) of CG4068, and an intergenic region adjacent to CG4770 (Supplementary Figs 2–7). Except for CG4068, the coding potential of all of these loci is limited^{5,6}. Still, we chose to introduce an 'hp' prefix to these six loci to distinguish the small-RNA-generating hairpins ('hpRNAs') from the potential protein-encoding segments of these transcripts.

The hpRNA hairpins were collectively much longer than typical animal pre-miRNAs, and several were even longer than plant miRNAs⁷. All hpRNA loci produced dominant small RNAs that pre-

sented duplexes with 2-nucleotide 3' overhangs, implying RNase III processing (Fig. 1 and Supplementary Figs 2–7). Such an origin was more evident with the hp-CG18854 and hp-CG4068 hairpins, from which many consecutive, phased, small RNA duplexes were cloned (Fig. 1). The 20 tandem repeats at the hp-CG4068 locus were suggestive of local duplications, and created potential for a vast array of higher-order hairpin conformations (Fig. 1b and Supplementary Fig. 3). In addition, hp-CR32205, hp-pncr009 and hp-CG32207 were related in sequence and located within a 70-kb interval (Supplementary Figs 5–8). Thus, hpRNAs, like miRNAs, can apparently evolve as local genomic duplications.

We probed the consequences of dsRNA-mediated knockdown of candidate factors on hpRNA biogenesis. We first confirmed the potency of these knockdowns by analysing Bantam, the pre-miRNA and/or mature miRNA of which were sensitive to Drosha, Pasha, Dcr-1, Loquacious (Loqs), Exportin-5 (Exp5, also known as Ranbp21) and Argonaute 1 (AGO1), as expected (Fig. 2a). The behaviour of hp-CG4068B/D/G and hp-CG18854A contrasted sharply with that of bantam (Fig. 2a and Supplementary Fig. 10a). Consistent with their apparent derivation from phased cleavage of long inverted repeats, their processing was unaffected by Dcr-1 depletion, but was strongly dependent on Dcr-2. In addition, ~21-nucleotide hpRNA products were markedly reduced when AGO2 was depleted. Incidentally, the Dcr-2/AGO2-dependent accumulation of ~21-nucleotide (siRNA) and ~42-nucleotide (terminal loop) hp-CG4068D isoforms (Fig. 2a) provided evidence for the *in vivo* processing of both 'single-repeat' and 'double-repeat' (or higher-order) forms of the hp-CG4068 hairpins (Fig. 1b).

Several other aspects of hpRNA biogenesis deserve mention. First, we were surprised that hpRNA processing was very strongly dependent on the Dcr-1-cofactor Loqs. This was especially unexpected in light of the recent realization that the *loqs* null condition only mildly compromises the maturation of many miRNAs⁸, such as Bantam (Fig. 2a). Second, mature hpRNA products declined reproducibly in AGO1-deficient cells, which suggested the possible involvement of both AGO proteins in hpRNA biogenesis and/or function. Third, knockdown of Dcr-2, AGO2 and, to a lesser extent, AGO1 resulted in a ladder of hybridizing bands consistent with impaired hairpin processing (Fig. 2a and Supplementary Fig. 10a). This suggested that, in addition to Dcr-2, AGO proteins might also participate in hpRNA biogenesis. A role for AGO proteins has also been suggested for the maturation of siRNA duplexes and some pre-miRNAs^{2,9–12}. Analysis of mutant animals corroborated this picture of hpRNA biogenesis, because mature hpRNA products were strongly reduced in *Dcr-2*, *loqs* and AGO2 homozygous mutants (Fig. 2b).

We next analysed the termini of hpRNA-derived small RNAs. β -elimination of RNAs with two free hydroxyl groups at their 3' termini increases their mobility in denaturing polyacrylamide gel

¹Sloan-Kettering Institute, Department of Developmental Biology, 521 Rockefeller Research Laboratories, 1275 York Avenue, Box 252, New York, New York 10065, USA. ²Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, and Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.

electrophoresis, whereas treatment with calf intestinal phosphatase (CIP) reduces the mobility of 5' monophosphorylated RNAs¹³. Accordingly, miRNAs run faster after β -elimination and slower after CIP treatment (Fig. 2c and Supplementary Fig. 10b). CIP tests also indicated the presence of 5' phosphates on hpRNA products (Supplementary Fig. 10b), but all of them were resistant to β -elimination indicating modification of the 3'-terminal ribose (Fig. 2c). *Drosophila* Hen1 methylates Piwi-associated RNAs (piRNAs) and exogenous siRNAs at their 3' termini^{14,15}. We found that *hen1* mutants exhibited lower levels of mature hpRNA products (Fig. 2b), and these were now fully sensitive to β -elimination

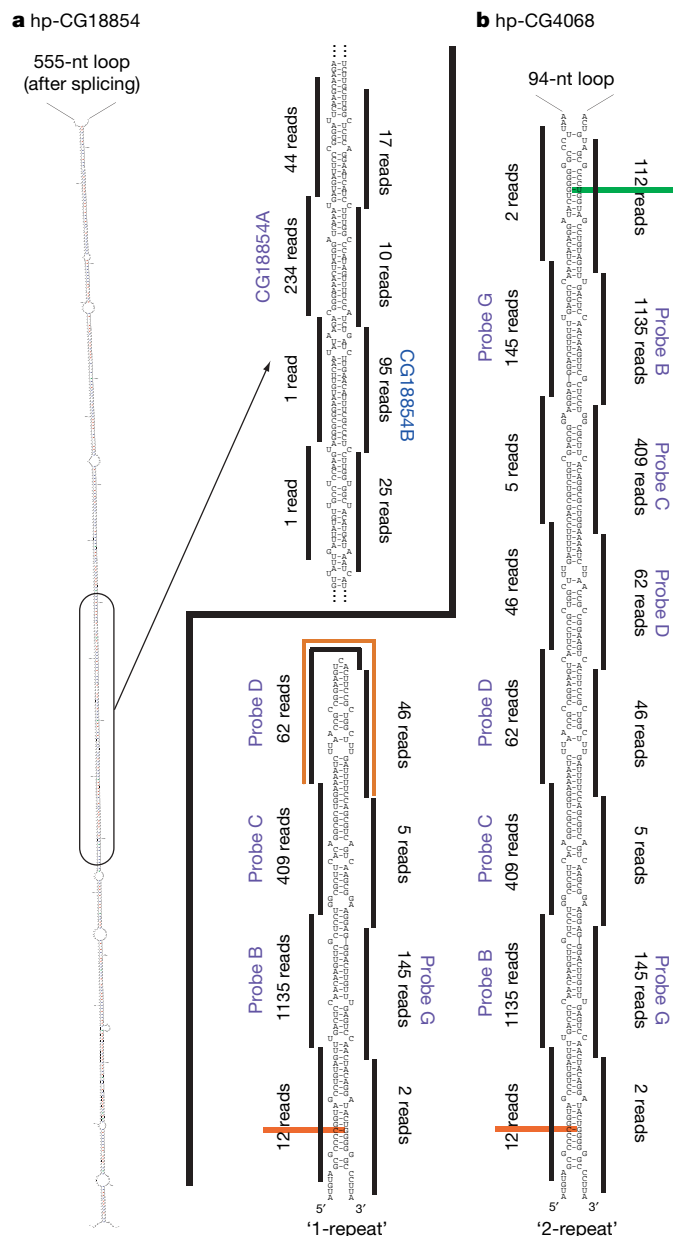


Figure 1 | Examples of *Drosophila* hpRNA transcripts. **a**, hp-CG18854 contains a >400-bp duplex separated by a large loop; the enlarged region highlights the phased nature of small RNA duplexes. Northern probes were designed against the RNAs labelled in blue. **b**, The hp-CG4068 locus consists of 20 tandem repeats that partially overlap the 3' UTR of CG4068 (Supplementary Fig. 3) and generate phased small RNA duplexes. Each repeat adopts a hairpin structure, but higher-order hairpins are possible because repeats are complementary to each other; '1-repeat' and '2-repeat' isoforms are shown. Distinct small RNAs were cloned from related repeats with minor sequence differences (for example, RNAs highlighted in red and green).

(Fig. 2c). These data supported the classification of hpRNA products not as miRNAs, but as siRNAs.

We tested the regulatory activity of hpRNA-derived siRNAs using artificial luciferase transcripts linked to target sites that were complementary to various hp-CG4068- and hp-CG18854-derived siRNAs. Their activity was analysed in cells that overexpressed hp-CG4068 or hp-CG18854, with non-cognate pairs controlling for the generic effect of hpRNA overexpression. These tests revealed the specific repression of hp-CG4068B and hp-CG4068C sensors by single- and double-repeat hp-CG4068 expression constructs (Fig. 3a), and of the hp-CG18854B sensor by hp-CG18854 (Fig. 3b). However, a sensor for hp-CG4068D was not affected by ectopic hp-CG4068, consistent with its lower read count compared to hp-CG4068B and hp-CG4068C.

To address the activity of endogenous hpRNAs expressed by S2 cells, we asked whether 2'-O-methyl antisense oligonucleotides (ASOs) could derepress these sensors. Indeed, ASO-hp-CG18854B (but not ASOs to hp-CG18854A or hp-CG4068B) induced approximately twofold derepression of the hp-CG18854B sensor (Fig. 3c). Reciprocally, we observed that ASO-hp-CG4068B (but not other ASOs) resulted in a approximately twofold activation of the hp-CG4068B sensor (Fig. 3d). Thus, both exogenous and endogenous hpRNAs generate inhibitory RNAs.

Some miRNAs are partially loaded into AGO2 (refs 16 and 17), but hpRNA products are the first endogenous *Drosophila* small RNAs known to be preferentially sorted to AGO2 as a class. This provided an opportunity to ask whether endogenously programmed AGO2 functions by means of slicing, translational repression, or both. We prepared hp-CG4068B sensors carrying tandem perfect sites, centrally bulged sites, or bulged plus seed mismatched sites. Both mutant sensors were strongly derepressed, and to roughly the same extent, relative to the perfect sensor (Fig. 3e). In fact, the activity of the

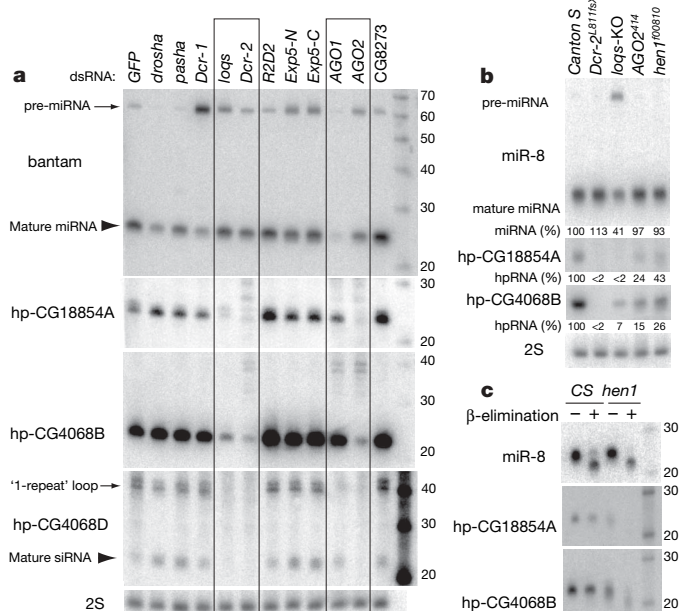


Figure 2 | Distinct biogenesis pathways for miRNAs and hpRNAs. **a**, Unlike miRNAs (for example, Bantam), hpRNA biogenesis in S2 cells is highly dependent on Dcr-2 and AGO2; Loqs and AGO1 suppression affect both miRNA and hpRNA biogenesis. **b**, miRNA and hpRNA biogenesis in pharate adult *Drosophila*. miR-8 was affected only in the *loqs* mutant, whereas hpRNA products were strongly decreased in the *Dcr-2* and *loqs* mutants (<10%), and significantly affected in the *AGO2* and *hen1* mutants. **c**, Modification of hpRNA-derived small RNAs is mediated by Hen1. Note that hpRNA products from *hen1* mutants run as a range of faster-migrating species after β -elimination. CS, Canton S (a strain of fruitfly); KO, knockout deletion strain.

mutant sensors was comparable to the perfect sensor in the presence of cognate ASO (Fig. 3d). These data support the notion that hp-CG4068B exerts its major regulatory effect by slicing, with comparably little contribution from translational repression by AGO2.

With these functional regulatory data in hand, we searched for endogenous targets. We first considered whether hp-CG4068 might regulate the overlapping CG4068 3' UTR (Supplementary Fig. 3), but gain- and loss-of-function tests were negative (Supplementary Fig. 11). Searches for *trans*-encoded targets were hindered by the fact that hp-CG4068 was only identifiable in the related species *D. melanogaster*, *Drosophila simulans* and *Drosophila sechellia* (Supplementary Fig. 3). However, hp-CG4068B, the most abundant siRNA product of hp-CG4068, contains 20 nucleotides of antisense complementarity (including three G•U pairs) to the coding region of *mus308* (Fig. 3f). This transcript encodes a DNA polymerase/helicase required for DNA repair after exposure to crosslinking agents¹⁸. Compensatory co-variation between *D. melanogaster* and *D. sechellia* hp-CG4068B and *mus308* target sites were suggestive of functional conservation (Fig. 3f). Consistent with this, we observed that *mus308* levels were increased ~2–3-fold in cells depleted of Dcr-2 or AGO2 (Supplementary Fig. 12), and that a luciferase–*mus308* sensor was specifically repressed by single-repeat and double-repeat hp-CG4068 expression constructs (Fig. 3a).

Because our data indicated that hpRNAs generate functional siRNAs that are primarily dependent on AGO2, we tested whether endogenous hp-CG4068B complexes exhibited Slicer activity. Endogenously programmed complexes indeed cleaved a perfect hp-CG4068B target substrate in a manner that was competed away by ASO-hp-CG4068B but not ASO-CG18854B (Fig. 3g). We also

found that a *mus308* target was cleaved by endogenous hp-CG4068B with similar specificity (Fig. 3g). We conclude that *mus308* is an endogenous Slicer target of hp-CG4068B.

We also searched for targets of hp-CG18854. The gene annotated as CG18854 is a possible pseudogene, because its open reading frame is short and poorly conserved⁵. CG18854 exhibits significant homology to the chromodomain gene CG8289 (Supplementary Fig. 13), and some of the abundant hp-CG18854-derived siRNAs exhibited extensive antisense complementarity to CG8289. When tested individually, siRNA-complementary sites from CG8289 did not mediate significant repression (data not shown). We therefore examined whether hp-CG18854 could regulate a translational fusion of CG8289 containing an extended complementary sequence. We transfected S2 cells with either tub-GFP or a tub-CG8289:GFP plasmid along with various hpRNA expression constructs, and observed that hp-CG18854 specifically repressed the accumulation of CG8289:GFP (Supplementary Fig. 13). These data demonstrate that hpRNA products can repress endogenous targets.

In plants, long hairpin RNAs from transgenes and long, extensively paired (and presumably very young) miRNA hairpins are substrates of DICER-LIKE4 (refs 19 and 20), and thus mature through a pathway distinct from that of canonical miRNA hairpins, which are substrates of DICER-LIKE1 (ref. 21). Likewise, we have shown that *Drosophila* hpRNAs enter a pathway distinct from that of miRNAs. Their derivation from unexpectedly long hairpins serves as an important caution for efforts to identify inverted-repeat small RNA genes. For example, some hpRNA-derived clones were recently reported but attributed incorrectly²², because only short genomic precursors were considered in that study.

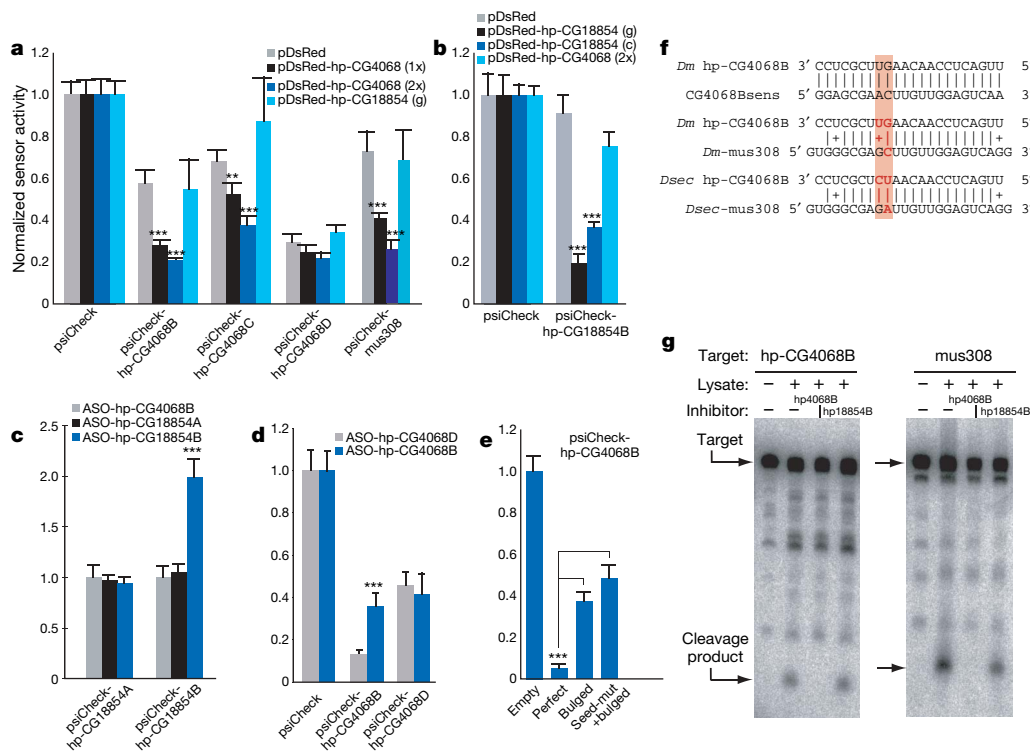


Figure 3 | hpRNAs generate regulatory RNAs that can repress endogenous targets. **a**, In S2 cells, both single-repeat (1×) and double-repeat (2×) hp-CG4068 constructs specifically repressed the hp-CG4068B, hp-CG4068C and mus308 sensors; the hp-CG4068D sensor was unaffected. **b**, Both genomic (g) and cDNA (c) hp-CG18854 expression constructs specifically repressed the hp-CG18854B sensor. **c**, The 2'-O-methyl ASO against hp-CG18854B specifically derepressed the hp-CG18854B sensor, whereas ASO-CG4068B specifically relieved endogenous repression of the hp-CG4068B sensor (**d**). **e**, Compared to a perfect hp-CG4068B sensor, mutant sensors

with a central bulge or central bulge plus seed mutations exhibited the same level of derepression. Error bars depict the standard deviation of eight transfections; statistical comparisons were performed with the unequal variance Students *t*-test; $^{*}P < 6 \times 10^{-5}$, $^{***}P < 1 \times 10^{-8}$. **f**, Compensatory covariation (pink shaded box) between hp-CG4068B and mus308 target sites of *D. melanogaster* (Dm) and *D. sechellia* (Dsec). Red font, nucleotides that have evolved; sens, sensor. **g**, Endogenous RNA-induced silencing complex from S2R cleaved both perfect hp-CG4068B and mus308 targets, and this activity was specifically competed away by cognate ASO.

The *Drosophila* pathway combines canonical RNAi (Dcr-2, Hen1 and AGO2) and miRNA (Loqs) biogenesis factors—a revelation that highlights the incomplete nature of our current understanding of small-RNA-sorting mechanisms. Together with concurrent studies that identify endogenous siRNAs from transposons and *cis*-natural antisense pairs in *Drosophila*^{22–24}, this work sets the stage for directed studies of the genetic requirements for host-directed RNAi in this organism.

METHODS SUMMARY

We used EINVERTED⁴ to identify candidate genomic hairpins contained within 10-kb windows that satisfied a cutoff score ≥ 80 and had $\geq 70\%$ pairing within the duplex region. Their expression as small RNAs was analysed using ten 454 libraries⁵, a Solexa female head library²⁵, and a new set of Solexa imaginal disc/brain library (NCBI-GEO GSM275691). We defined candidate hpRNA loci as non-transposon inverted repeats for which the duplex region generated more than three times as many 21–22-nucleotide RNAs than all other-sized RNAs combined (Supplementary Figs 2–9). For functional tests, we followed published protocols for soaking RNAi in S2 cells and northern blotting²⁵ from knockdown samples or pharate adult flies. For sensor tests, four-copy-site targets (hp-CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (*mus308* sensor) were prepared by inserting oligonucleotides into a modified version of psiCHECK2 (ref. 25). For hpRNA expression constructs, one or two hp-CG4068 repeats were cloned into the 3' UTR of UAS-DsRed; CG18854 fragments were amplified from genomic DNA (CG18854 genomic) or a LD34273 clone (CG18854 cDNA) and were cloned similarly. A CG8289:GFP translational sensor consisted of a CG8289 fragment from genomic DNA and cloned into the KpnI site of tub-GFP²⁶. RNA 3' termini were analysed using periodate treatment in borax/boric-acid buffer followed by NaOH treatment (β -elimination) as described¹⁵. RNA 5' termini were analysed using CIP as described¹³. For cleavage assays, we prepared labelled gel-purified templates using α -³²P-GTP and capping enzyme (Ambion). Cleavage reactions were performed as described² using S2-R+ cell extract. For detailed bioinformatic and molecular methods, see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 22 April 2008.

Published online 7 May 2008.

1. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
2. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
3. Kennerdell, J. R. & Carthew, R. W. Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nature Biotechnol.* **18**, 896–898 (2000).
4. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
5. Wilson, R. J., Goodman, J. L. & Strelets, V. B. FlyBase: integration and improvements to query tools. *Nucleic Acids Res.* **36**, D588–D593 (2008).
6. Tupy, J. L. *et al.* Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **102**, 5495–5500 (2005).
7. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53 (2006).
8. Liu, X. *et al.* Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13**, 2324–2329 (2007).
9. Leuschner, P. J., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).

10. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
11. Diederichs, S. & Haber, D. A. Dual role for Argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
12. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
13. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
14. Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* **21**, 1603–1608 (2007).
15. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
16. Seitz, H., Ghildiyal, M. & Zamore, P. D. Argonaute loading improves the 5' precision of both microRNAs and their miRNA strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
17. Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
18. Harris, P. V. *et al.* Molecular cloning of *Drosophila mus308*, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. *Mol. Cell. Biol.* **16**, 5764–5771 (1996).
19. Dunoyer, P., Himber, C. & Voinnet, O. DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nature Genet.* **37**, 1356–1360 (2005).
20. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**, 3407–3425 (2006).
21. Reinhardt, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
22. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* doi:10.1126/science.1157396; published online 10 April 2008.
23. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* doi:10.1038/nature07007 (this issue).
24. Kawamura, Y. *et al.* *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* doi:10.1038/nature06938 (this issue).
25. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100 (2007).
26. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of *Drosophila* microRNA targets. *PLoS Biol.* **1**, E60 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to R. Carthew, Q. Liu, H. Siomi and S. Cohen for plasmids and *Drosophila* strains. K.O. was supported by the Charles H. Revson Foundation. H.G. was supported by A*STAR, Singapore. D.P.B. is an HHMI investigator, and work in his laboratory was supported by a grant from the NIH (GM067031). E.C.L. was supported by grants from the Leukemia and Lymphoma Foundation, the Burroughs Wellcome Foundation, the V Foundation for Cancer Research, the Sidney Kimmel Foundation for Cancer Research, and the NIH (GM083300).

Author Contributions J.G.R. identified hp-CG4068 and hpRNA1. W.-J.C. performed the EINVERTED analysis and identified the additional hpRNA loci and their targets. H.G. performed initial hpRNA northern analysis; all other experiments were designed and carried out by K.O. All authors contributed to the preparation of the manuscript.

Author Information The imaginal disc/brain sample described in paper has been deposited in the NCBI GEO under accession number GSM275691. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.C.L. (laie@mskcc.org).

METHODS

Bioinformatics. We used EINVERTED⁴ to identify candidate genomic hairpins contained within 10-kb windows that satisfied a cutoff score ≥ 80 and had $\geq 70\%$ pairing within the duplex region. These criteria eliminated all but one of the annotated *Drosophila* miRNAs (*mir-997*)²⁷. We kept those candidates with small RNA reads in the following data sets: 10 libraries analysed using 454 pyrosequencing⁵, a female head library analysed using Solexa²³, and a new set of imaginal disc/brain small RNA sequences analysed using Solexa. We removed those loci in which the predicted duplex overlapped an annotated transposable element²⁸, and calculated the size distribution of reads from each of the remaining loci. We considered those loci for which the duplex region generated more than three times as many 21–22-nucleotide RNAs than all other-sized RNAs combined as hpRNA candidates (Supplementary Figs 2–9).

RNA interference. Segments of Pasha, Drosha and CG8273 were amplified using the indicated primers and were cloned into XhoI–XbaI sites of Litmus28i vector (NEB); other plasmids were described previously^{25,29}. These templates were used to generate dsRNA, and soaking RNAi was performed as described²⁵. S2-R+ cells were resuspended in serum-free medium at 3×10^6 cells ml⁻¹ density and dsRNA was added to a concentration of 15 μ g ml⁻¹. After 30 min incubation, an equal volume of Schneider's medium supplemented with 20% FBS was added. dsRNA treatment was repeated 4 days after the first treatment, and RNA samples were collected 4 days after the second soaking. XhoI-CG8273-279+, AGAGC-TCGAGTCAGACAAATCCTCCGGTTC; XbaI-CG8273-682–, AGAGGTCTA-GATTCCGATCTGACTTGGTTC; XhoI-Pasha-1240+, AGAGCTCGAGGG-AGGTGGAGCAACAAAAGA; XbaI-Pasha-1725–, AGAGTCTAGATACCTC-GTGCAGGATGCAGAC; XhoI-Drosha3522+, AGAGCTCGAGGCCGAGACA-TTCCCTACTACA; XbaI-Drosha3943–, AGAGGTCTAGAGGCCATTGTT-GGACTCTTG.

Northern blotting. Northern blotting was performed as described²⁵ using total RNA isolated from S2 cells or pharate adult *Drosophila*. All of the mutant strains used were described previously: *Dcr-2*^{L811f5X} (ref. 1), *logq*-KO (ref. 30), *AGO2*⁴¹⁴ (ref. 2) and *hen1*¹⁰⁰⁸¹⁰ (refs 14 and 15). The sequences of the probes used in this study are listed below. DNA and locked nucleic acid (LNA) probes were obtained from IDT and Exiqon, respectively. *bantam* probe (DNA), AATCAGCTTT-CAAAATGATCTCA; 2S rRNA probe (DNA), TACAACCTCAACCATA-TGTAGTCCAAGCA; hp-CG4068-D (LNA), GTGACTTCCGGCGGTTAA-GATTT; hp-CG4068-B (LNA), GGAGCGAAGTGTGGAGTCAA; hp-CG4068-G (LNA), AGTTGGACTCAACAAGTCCCT; hp-CG18854-A (LNA), TCATTGTATCCATAGTTTCCCGT; hp-CG18854-B (LNA), GGAGG-CGGAATGTTCAAGATCA; miR-8 (LNA), A.

Analysis of RNA chemical structure. RNA 3' termini were analysed as described¹⁵. 10 μ l of 61.5 mM NaIO₄ in borax/boric-acid buffer (60 mM borax and 60 mM boric acid, pH 8.6) was added to 10 μ g total RNA in 14.6 μ l water, and the samples were incubated for 30 min at room temperature (22 °C). 2.5 μ l of 500 mM NaOH was added to each sample and incubation was continued for 90 min at 45 °C. The reactions were stopped by addition of 200 μ l of 300 mM NaCl, 10 μ g glycogen and 600 μ l absolute ethanol. RNA was collected by centrifugation after 30 min incubation on ice.

RNA 5' termini were analysed as described¹³. RNA samples were incubated with 2 units CIP (New England Biolabs) in 1 \times buffer 3 (NEB, 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM dithiothreitol) for 2 h at 37 °C. RNA was purified by phenol/chloroform extraction followed by ethanol precipitation.

Sensor assays. Four-copy-site targets (hp-CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (*mus308* sensor) were prepared using the oligonucleotides listed below. Target sequences were inserted into SalI–XhoI (four-copy sensors) or NotI–XhoI (two-copy sensor) cloning sites of a modified version of psiCHECK2 (ref. 25). CG4068 3' UTR sensor was constructed by inserting a CG4068 3' UTR fragment amplified with primers CG4068A and B (containing one repeat of the hpRNA repeat) into NotI–XhoI sites of the modified psiCHECK2. CG18854 fragments were amplified from genomic DNA (CG18854 genomic) or LD34273 clone (CG18854 cDNA) and cloned into NotI–XhoI sites of UAS–DsRed. The CG8289 fragment was amplified from genomic DNA and cloned into KpnI site of tub-GFP plasmid²⁶.

Luciferase sensor assays were performed as described previously²⁵. We performed quadruplicate transfections of 25 ng target, 12.5 ng ub-Gal4 and 25 ng UAS–DsRed–hpRNA plasmids into 1×10^5 S2 cells in 96-well format. For 2'-O-methyl antisense-mediated de-silencing assays (inhibitor sequences listed below), we introduced 25 ng target plasmid and 10 pmol of 2'-O-methyl oligonucleotides for each well. Three days later, we lysed the cells and subjected them to the dual luciferase assay (Promega) and analysed these on a Veritas plate luminometer (Turner Biosystems). KpnI-CG8289 targetF, ggggtaccgcccacc-

atgTTGCTGAAAAGGATTTCG; KpnI-CG8289, ggggtaccTTCGAGGAGCG-TTCAATACGAT targetR; NotI-LD34273-1+, AGAGcggcgAGTGCTG-AGCATACCTAAGC; XhoI-LD34273-2390–, AGAGctcgagGTTCCCATC-TGACTGGAAT; CG4068_A, agggcgccgcACAAGCAAAATCGTAtagg; CG4068_B, agggctcgagTTTTGCGTGGACTCATTTCC; hp-CG4068B_targ_A, tcgacaaaaGGAGCGAAGTGTGGAGTCAAagaac; hp-CG4068B_targ_B, tcgagttctTTGACTCCAACAAGTTCGCTCCTttt; hp-CG4068C_targ_A, tcgacaaaaTTCCAGCGCTGTGAAGCGCCAgagaac; hp-CG4068C_targ_B, tcgagttctTGGCGCTTCACAGGCGCTGGAAtttt; hp-CG4068D_targ_A, tcgacaaaaGTGACTTCCGCGGTTAAGATTTagaac; hp-CG4068D_targ_B, tcgagttctAAATCTTAACCGCCGGAAGTCACTttt; hp-CG4068A_si2x_A, GGCCGCGGAGCGAAGTGTGGAGTCAAaaatcacGGAGCGAAGTGTGTG-GAGTCAAaC; hp-CG4068A_si2x_B, TCGAGTTCGACTCCAACAAGTTC-GTCCGtgattTTGACTCCAACAAGTTCGCT GC; hp-CG4068A_mi2x_A, GGCCGCGGAGCGCAACTACATCCACTGAaaatcacGGAGCGAAGTACATC-CACTGAAaC; hp-CG4068A_mi2x_B, TCGAGTTCAGTGGATGTAGTTCG-CTCCGtgattTTGAGTGGATGTAGTTCGCT GC; hp-CG4068A_mimut2x_A, GGCCGCGGAGCGCAACTACATGGAGTCAAaaatcacGGAGCGAAGTACATG-GAGTCAAaC; hp-CG4068A_mimut2x_B, TCGAGTTCGACTCCATGTAGTT-CGCTCCGtgattTTGACTCCATGTAGTTCGCT GC; hp-CG4068A_mimut2x_B, gggcattggCGAGCTTGTGGAGTCAAggttgattggGCGAGCTTGTGGAGTC-Aggc; mus308-target2, tcgagcctgactcaacaagctgcccaatcacccTGACTCCAACA-AGTTCGCCat; ASO-hp-CG4068-B, 5' AACAUggagcgaactgttgtagtcaaUACA 3'; ASO-hp-CG4068-C, 5' AACAUttcagcgctgtgaagcgccagUACACU 3'; ASO-hp-CG4068-D, 5' AACAUUGACUUCGCGCGGUUAGAUAUUAUACA 3'; ASO-hp-CG18854-A, 5' AACAUtgGCCAAGGTACGTGGTTCGACGAUAUA-CU 3'; ASO-hp-CG18854-B, 5' AACAUUGAGGGCGAAATGTTCAAGATC-AUCACU 3'.

For the green fluorescent protein (GFP) sensor assay, 250 ng target, 125 ng ub-Gal4 and 250 ng UAS–DsRed–hpRNA plasmids were transfected to 2×10^6 cells in 6-well format. Three days later, transfected cells were harvested and lysed with 2 \times SDS–PAGE sample buffer. Western blotting was performed using rabbit anti-GFP (Molecular Probes) or mouse anti- α -tubulin (DM1A, Sigma).

In vitro cleavage assay. Templates for *in vitro* cleavage targets were prepared by treating annealed oligonucleotides with Taq polymerase. The oligonucleotides A and B or A and C (below) were used for the template preparation for hp-CG4068-B target or mus308 target, respectively. Target RNAs were *in vitro* transcribed using Megascript T7 kit (Ambion) and purified by acrylamide gel electrophoresis. Purified RNA was labelled by α -³²P-GTP using capping enzyme (Ambion) according to the manufacturer's instructions. The cleavage reaction was performed as described² using S2-R+ cell extract. S2-R+ cells were resuspended in hypotonic buffer (30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate, 5 mM DTT, 1 \times Complete mini EDTA free (Roche)) and lysed by five passages through a 25-gauge needle. The lysate was cleared by a centrifugation for 25 min at 14,600g at 4 °C, and was flash-frozen in 10- μ l aliquots. Approximately 2,000 counts per min cap-labelled RNA was incubated in a reaction mixture (50% S2 lysate, 0.5 mM ATP, 5 mM DTT, 100 mM KOAc, 0.1 U μ l⁻¹ RNaseOut (Invitrogen)) for 1 h at room temperature. 2'-O-methyl-ASO inhibitors (Integrated DNA Technologies) were added at 100 nM concentration and preincubated with the reaction mixture at room temperature for 20 min before the addition of the cap-labelled target RNA. Reactions were stopped by the addition of stop buffer (50 mM sodium chloride, 50 mM EDTA, 1% SDS, and 100 μ g ml⁻¹ proteinase K). RNA was recovered by phenol/chloroform extraction and ethanol precipitation. A, LucLet7_3' region_AS, attaatcctatGAGGTAGTAGTTGTATAGTTCGAAGTATTCGCGTACGTT; B, T7_hp-CG40682B_Luc_sense, TAATACGACTCACTATAGGAGCGAAC-TTGTTGGAGTCAAattAACGTACGCGGAATAC; C, T7_Mus308tar_Luc_sense, TAATACGACTCACTATAGgtggGCGAGCTTGTGGAGTCAgggAACGTACGCGGAATAC.

27. Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* 17, 1850–1864 (2007).
28. Karolchik, D. *et al.* The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* 36, D773–D779 (2008).
29. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* 3, e236 (2005).
30. Park, J. K., Liu, X., Strauss, T. J., McKearin, D. M. & Liu, Q. The miRNA pathway intrinsically controls self-renewal of *Drosophila* germline stem cells. *Curr. Biol.* 17, 533–538 (2007).

NF- κ B links innate immunity to the hypoxic response through transcriptional regulation of HIF-1 α

Jordi Rius^{1,2,3}, Monica Guma^{1,2,3}, Christian Schachtrup², Katerina Akassoglou², Annelies S. Zinkernagel⁴, Victor Nizet^{4,5}, Randall S. Johnson⁶, Gabriel G. Haddad⁴ & Michael Karin^{1,2,3}

The hypoxic response is an ancient stress response triggered by low ambient oxygen (O₂) (ref. 1) and controlled by hypoxia-inducible transcription factor-1 (HIF-1), whose α subunit is rapidly degraded under normoxia but stabilized when O₂-dependent prolyl hydroxylases (PHDs) that target its O₂-dependent degradation domain are inhibited^{2–4}. Thus, the amount of HIF-1 α , which controls genes involved in energy metabolism and angiogenesis, is regulated post-translationally. Another ancient stress response is the innate immune response, regulated by several transcription factors, among which NF- κ B plays a central role^{5,6}. NF- κ B activation is controlled by I κ B kinases (IKK), mainly IKK- β , needed for phosphorylation-induced degradation of I κ B inhibitors in response to infection and inflammation⁷. IKK- β is modestly activated in hypoxic cell cultures when PHDs that attenuate its activation are inhibited⁸. However, defining the relationship between NF- κ B and HIF-1 α has proven elusive. Using *in vitro* systems, it was reported that HIF-1 α activates NF- κ B⁹, that NF- κ B controls HIF-1 α transcription¹⁰ and that HIF-1 α activation may be concurrent with inhibition of NF- κ B¹¹. Here we show, with the use of mice lacking IKK- β in different cell types, that NF- κ B is a critical transcriptional activator of HIF-1 α and that basal NF- κ B activity is required for HIF-1 α protein accumulation under hypoxia in cultured cells and in the liver and brain of hypoxic animals. IKK- β deficiency results in defective induction of HIF-1 α target genes including vascular endothelial growth factor. IKK- β is also essential for HIF-1 α accumulation in macrophages experiencing a bacterial infection. Hence, IKK- β is an important physiological contributor to the hypoxic response, linking it to innate immunity and inflammation.

Hypoxia is characterized by a decreased O₂ tension within cells and can occur under several pathophysiological situations including ischaemia, cancer and inflammation¹². During ischaemia, the flow of nutrients and O₂ to damaged tissues is decreased and HIF-1 α activation induces genes whose products restore blood supply, nutrients and energy production, thereby maintaining tissue integrity and homeostasis^{13,14}. The hypoxic response is important for the proper function of tissue macrophages and infiltrating neutrophils that encounter low O₂ tension in infected tissues and after bacterial replication¹⁵. HIF-1 α was also suggested to promote the expression of inflammatory cytokines, which are known to be regulated by NF- κ B¹⁶, in lipopolysaccharide (LPS)-stimulated macrophages¹⁷ and mediate NF- κ B activation in anoxic neutrophils⁹. However, it was also reported that hypoxia leads to modest IKK- β activation by inhibiting PHDs that negatively modulate IKK- β activity⁸. We therefore decided to critically explore the relationship between IKK- β , NF- κ B and HIF-1 α under *in vivo* conditions in IKK- β -deficient mice and primary macrophages.

We first examined bone marrow-derived macrophages (BMDM) from either IKK $\beta^{F/F}$ or IKK $\beta^{F/F}/Mx1Cre$ mice challenged with poly(I)•poly(C), which induces interferon (IFN) and thereby drives CRE recombinase expression from the *Mx1* promoter to delete IKK β in IFN-responsive cells of the resulting IKK β^A mice¹⁸. BMDM were incubated with Gram-positive (group A *Streptococcus*; GAS) and with Gram-negative (*Pseudomonas aeruginosa*) bacteria. Both species induced HIF-1 α accumulation in an IKK- β -dependent manner (Fig. 1a). The induction of HIF-1 target genes involved in the hypoxic and innate immune responses was also dependent on IKK- β (Fig. 1b). These genes included *Cox-2*, which is directly regulated by NF- κ B and HIF-1 α , *Cnlp*, which encodes the murine antimicrobial peptide mCRAMP, whose expression is not directly responsive to NF- κ B¹⁹, and *Glut-1*, encoding a glucose transporter. Moreover, *Hif1a* mRNA was markedly downregulated in IKK- β -deficient cells even before infection (Fig. 1b). I κ B degradation and the nuclear accumulation of RelA/NF- κ B preceded HIF-1 α expression (Fig. 1c), indicating that NF- κ B may control *Hif1a* gene transcription. Indeed, chromatin immunoprecipitation (ChIP) in LPS-stimulated macrophages revealed that RelA is recruited to the *Hif1a* promoter, which contains a classical κ B site at –197/–188 base pairs, conserved between mice and humans (Fig. 1d). Furthermore, the basal levels of *Hif1a* mRNA were decreased in RelA-deficient fibroblasts even under resting conditions (Supplementary Fig. 1), suggesting that NF- κ B activity is required for effective *Hif1a* mRNA expression even in non-stimulated cells.

As found elsewhere⁸, hypoxia modestly activated IKK in macrophages (Fig. 2a), induced the phosphorylation of IKK- α/β and I κ B α and promoted I κ B α degradation (Fig. 2b). Hypoxia also induced the nuclear translocation of RelA, which preceded HIF-1 α accumulation (Fig. 2c), as occurred in bacteria-infected macrophages (Fig. 1c). Binding of NF- κ B to a canonical κ B DNA site was also induced by hypoxia (Fig. 2d). We examined whether IKK- β was required for hypoxia-induced HIF-1 α accumulation, a response that is thought to be dependent mainly on inhibition of HIF-1 α degradation^{3,4}. IKK- β was required for the optimal accumulation of HIF-1 α , but not of HIF-2 α , in BMDM incubated with the hypoxia mimetic desferrioxamine (DFX) as well as in response to actual hypoxia (Fig. 3a, b). IKK- β also did not affect HIF-2 α expression in infected macrophages (Fig. 1a). The overexpression of a non-degradable I κ B α (I κ B super-repressor) also blocked HIF-1 α accumulation induced by hypoxia in HEK-293 cells (Supplementary Fig. 2). The hypoxia-dependent induction of HIF-1 target genes, such as those encoding vascular endothelial growth factor (VEGF) and GLUT-1, was nearly abolished in IKK- β -deficient macrophages (Fig. 3c) or fibroblasts (Supplementary Fig. 3). Expression of *Hif1a*, but not *Hif2a*, mRNA was

¹Laboratory of Gene Regulation and Signal Transduction, ²Department of Pharmacology, ³Department of Pathology, ⁴Department of Pediatrics, School of Medicine, ⁵Skaggs School of Pharmacy and Pharmaceutical Sciences, and ⁶Molecular Biology Section, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0723, USA.

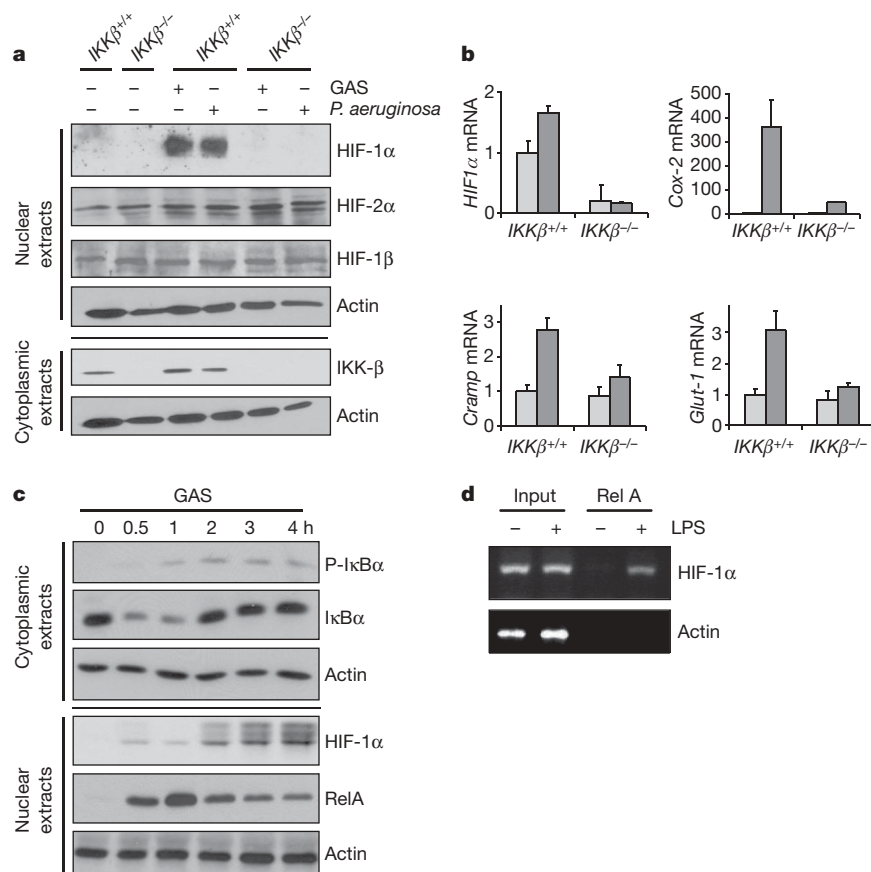


Figure 1 | IKK- β is required for microbial-induced HIF-1 α expression in macrophages. **a**, BMDM from either $IKK\beta^{F/F}$ ($IKK\beta^{+/+}$) or poly(I)•poly(C)-injected $IKK\beta^{F/F}$ /Mx1-Cre ($IKK\beta^{-/-}$; $IKK\beta^{+/+}$) mice were incubated with either with GAS or *P. aeruginosa* (MOI of 10 for 4 h). Expression of the indicated proteins was analysed by immunoblotting. **b**, RNA was extracted from BMDM incubated with GAS and gene expression was analysed by quantitative RT-PCR. Results are averages of three separate experiments done in triplicate, and are shown as means and s.e.m. Values are

normalized relative to 18S rRNA. Light grey bars, control; dark grey bars, GAS. *Cramp*, gene encoding cathelicidin antimicrobial peptide. **c**, RAW264.7 macrophages were incubated with GAS and protein expression was analysed by immunoblotting at the indicated time points. **d**, ChIP was performed with an anti-RelA antibody using fixed and sheared chromatin isolated from RAW264.7 mouse macrophages incubated for 1 h with or without LPS. The *Hif1a* promoter fragment, which contains a κ B site at -197/-188 base pairs, was detected by PCR amplification.

substantially decreased in the absence of IKK- β even under normoxia (Fig. 3c), further supporting the notion that basal NF- κ B activity is required for the expression of enough *Hif1a* mRNA at all times to result in the rapid accumulation of HIF-1 α protein, which occurs only under hypoxic conditions. Activation of NF- κ B by LPS induced *Hif1a* promoter activity (Supplementary Fig. 4), elevated HIF-1 α expression in hypoxic cells (Fig. 3d) and potentiated the induction of *Vegf* mRNA (Supplementary Fig. 5). Despite substantial expression of *Hif1a* mRNA in LPS-stimulated normoxic macrophages (Supplementary Fig. 5), these cells do not accumulate HIF-1 α protein (Fig. 3d), which echoes findings in T cells stimulated with anti-CD3 antibody²⁰. Hence, NF- κ B activation without hypoxic inhibition of PHDs is insufficient for HIF-1 α protein accumulation. In mouse fibroblasts, IKK- β was required for basal *Hif1a* promoter activity and its stimulation by treatment with DFX (Fig. 3e).

We next examined the role of IKK- β in HIF-1 activation in intact mice. Administration of DFX induced HIF-1 α expression in liver of $IKK\beta^{F/F}$ mice but not in $IKK\beta^{-/-}$ mice (Fig. 4a), which lack IKK- β in both hepatocytes and Kupffer cells²¹. $IKK\beta^{-/-}$ mice also contained less *Hif1a* and *Vegf* mRNA in their livers (Fig. 4b). We also examined the role of IKK- β in the response to actual hypoxia. Mice were placed in a chamber with an ambient O₂ concentration of 8% (thus mimicking an altitude of 7,000 m (ref. 22)). Under these conditions we observed hypoxia-induced HIF-1 α accumulation in liver (Fig. 4c) and brain (Fig. 4d) and in both cases it was dependent on IKK- β in CRE-expressing cells. In the brain the predominant CRE-expressing cells

were astrocytes (Supplementary Fig. 6) and not neurons (data not shown), thus explaining the partial deletion of IKK- β in this tissue (Fig. 4d). Despite this, hypoxia-induced VEGF protein (Fig. 4e) and *Vegf* mRNA (Fig. 4f) were IKK- β dependent. $IKK\beta^{-/-}$ mice showed a profound increase in cerebellar astrocyte activation, marked by glial fibrillary acidic protein, relative to $IKK\beta^{F/F}$ mice (Supplementary Fig. 7). This may have been due to defective production of VEGF, a cytokine with anti-inflammatory properties that has been shown to promote tissue repair²³. VEGF is also a potent neuroprotective factor²⁴ whose decreased production may potentiate hypoxia-induced neuronal damage and thereby augment astrocyte activation. This situation may be akin to a loss of IKK- β in intestinal epithelial cells, which has previously been found to exacerbate ischaemic damage to the intestinal mucosa²⁵. These results suggest that IKK- β inhibitors may not be useful in the treatment of neuroinflammatory disorders.

Although early studies demonstrated the induction of *Hif1a* mRNA in experimental animals during development and hypoxia^{26,27}, numerous *in vitro* studies led to the current model that the accumulation of HIF-1 α is regulated predominantly at the post-translational level through the inhibition of O₂-dependent PHDs that drive HIF-1 α degradation under normoxia^{3,4}. Our results show clearly that transcriptional activation of the *Hif1a* gene by IKK- β -responsive NF- κ B, which precedes HIF-1 α protein accumulation, is of critical importance under pathophysiologically relevant conditions *ex vivo* and *in vivo*. Both macrophages infected with bacteria and mice subjected to hypoxia reveal a pronounced defect in HIF-1 α expression on loss of

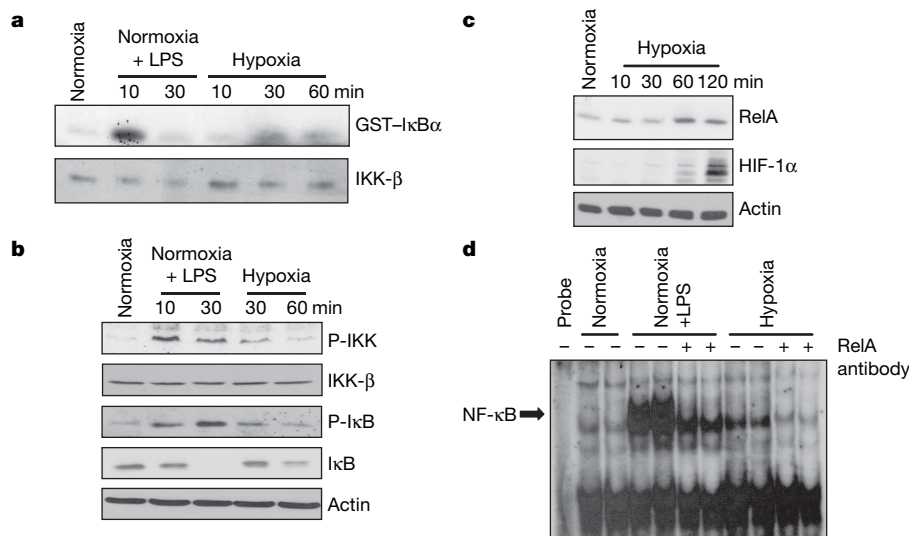


Figure 2 | Hypoxia activates the NF-κB pathway in macrophages.

RAW264.7 macrophages were incubated with or without LPS under normoxia or were placed under hypoxia (0.5% O₂). **a**, At the indicated time points, IKK activity was measured by an immunocomplex kinase assay with GST-IκBα as a substrate. **b**, Cell lysates were prepared and IKK-β and IκBα phosphorylation (P) and amounts were analysed by immunoblotting.

c, Nuclear extracts were prepared at the indicated time points and analysed by immunoblotting for the nuclear accumulation of RelA and HIF-1α.

d, Nuclear extracts were prepared after 2 h of stimulation with LPS or hypoxia, and binding activity for DNA for NF-κB was examined by a mobility-shift assay. Antibody inhibition was performed with an anti-RelA antibody.

IKK-β. The IKK-β/NF-κB-HIF-1α crosstalk is not critical during normal embryonic development, because the respective gene deletions result in different phenotypes. Whereas *Hif1a*^{-/-} embryos die prematurely at embryonic day 9.5, mainly as a result of defects in neural fold closure and capillary development^{13,14}, *IKK-β*^{-/-} embryos

die later, at embryonic day 13.5, from massive liver apoptosis driven by TNF^{28,29}.

Previous findings identified a connection between HIF-1α and innate immunity and inflammation, but it was not clear how microbial infection or inflammation led to HIF-1α activation^{15,19}. Our

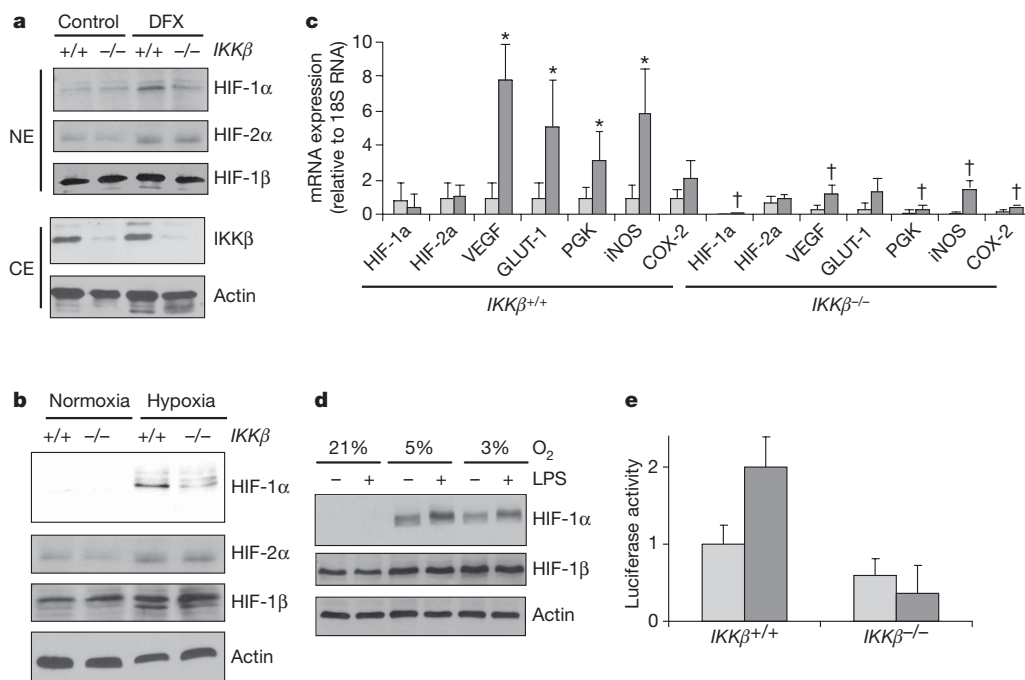


Figure 3 | IKK-β regulates hypoxia-induced HIF-1α and target genes in mouse macrophages.

a, BMDM from *IKKβ*^{F/F} (*IKKβ*^{+/+}) or *IKKβ*^d (*IKKβ*^{-/-}) mice were incubated with DFX for 4 h. Expression of the indicated proteins was analysed by immunoblotting for nuclear (NE) and cytosolic (CE) extracts. **b**, BMDM were obtained as above and cultured under normoxia or hypoxia (0.5% O₂ for 4 h). Protein expression was analysed by immunoblotting. **c**, BMDM were treated as above and mRNA expression was analysed by quantitative RT-PCR. Light grey bars, normoxia; dark grey bars, hypoxia. Results are means and s.e.m. for three separate experiments performed in triplicate. *, *P* < 0.05 versus normoxic *IKKβ*^{+/+} cells; †,

P < 0.05 versus hypoxic *IKKβ*^{+/+} cells. PGK, phosphoglucokinase; iNOS, inducible nitric oxide synthase. **d**, RAW264.7 macrophages were cultured in the absence or presence of LPS under the indicated O₂ tensions for 2 h. Protein expression was analysed by immunoblotting. **e**, Murine embryonic fibroblasts from *IKKβ*^{+/+} or *IKKβ*^{-/-} embryos were transfected with a luciferase reporter gene driven by the *Hif1a* promoter. After 36 h the cells were incubated for 3 h with DFX. Light grey bars, control; dark grey bars, DFX. Results are means and s.e.m. for three separate experiments performed in triplicate.

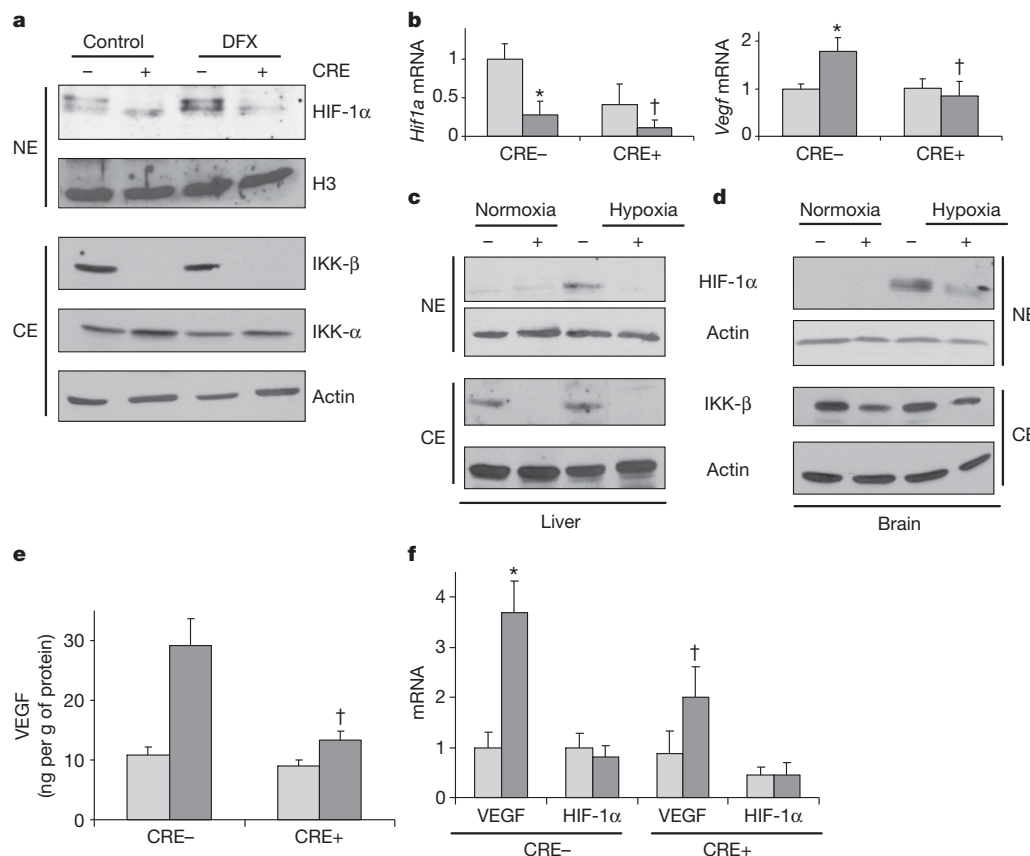


Figure 4 | **IKK- β regulates HIF-1 α expression in hypoxic mice.** *IKK β ^{F/F}* (CRE⁻) or *IKK β ^{F/F}* (CRE⁺) mice were treated with vehicle (control) or DFX (600 mg kg⁻¹). After 15 h, livers were removed for protein (a) and RNA (b) analysis. a, HIF-1 α and IKK- β expression in nuclear (NE) or cytosolic (CE) extracts was analysed by immunoblotting. b, Expression of HIF-1 α and *Vegf* mRNA was examined by quantitative RT-PCR. Light grey bars, control; dark grey bars, DFX. Results are means and s.e.m. ($n = 3$). Values are normalized relative to 18S rRNA. *, $P < 0.05$ versus normoxic CRE⁻ mice; †, $P < 0.05$ versus DFX-treated CRE⁻ mice. c, d, *IKK β ^{F/F}* and *IKK β ^{F/F}* mice were kept under normoxia or hypoxia (8% O₂) for 24 h and HIF-1 α and IKK-

β expression was analysed by immunoblotting of liver (c) or brain (d) nuclear and cytosolic extracts, respectively. e, VEGF expression in brain of mice from the above experiment was analysed by ELISA. Light grey bars, normoxia; dark grey bars, hypoxia. Results are means and s.e.m. ($n = 3$). *, $P < 0.05$ versus normoxic CRE⁻ mice; †, $P < 0.05$ versus hypoxic CRE⁻ mice. f, VEGF and *Hif1a* mRNA expression were analysed by quantitative RT-PCR of total brain RNA. Light grey bars, normoxia; dark grey bars, hypoxia. Results are means and s.e.m. *, $P < 0.05$ versus normoxic CRE⁻ mice; †, $P < 0.05$ versus hypoxic CRE⁻ mice ($n = 3$).

results, together with the previous finding that IKK- β catalytic activity is controlled by O₂-sensitive PHDs⁸, establish NF- κ B as a hypoxia-regulated transcription factor that controls *Hif1a* mRNA expression both under basal conditions and during hypoxia, thereby serving as a regulator of the hypoxic response. Our findings demonstrate that this depends on NF- κ B activation, which controls *Hif1a* mRNA expression, but accumulation of HIF-1 α protein requires hypoxia, which in bacterial infection may be due to depletion of intracellular oxygen by replicating bacteria. These findings have far-reaching physiological implications because they indicate the existence of coupling between two evolutionary ancient stress responses: innate immunity and the hypoxic response. By controlling HIF-1 α activation in macrophages during microbial infections, which may lower local O₂ tension, NF- κ B can enhance glycolytic energy metabolism and the production of angiogenic factors, in addition to its well-established role in the expression of proinflammatory cytokines, chemokines and antimicrobial peptides. In addition to more effective execution of the host-defence response, the ability of NF- κ B to promote HIF-1 α activation expands its pro-survival function because the HIF-1-dependent hypoxic response is critical for providing cells and tissues undergoing ischaemia with sufficient energy supplies and allows them to resist cell death.

By serving as an essential component of the hypoxic response *in vivo*, IKK- β also performs a homeostatic function in the brain,

an organ that is extremely sensitive to deprivation of oxygen and glucose³⁰.

METHODS SUMMARY

To delete IKK- β in *IKK β ^{F/F}/Mx1Cre* mice, 250 μ g of poly(I)•poly(C) (Sigma) was injected intraperitoneally on three alternate days, three weeks before exposure to hypoxia or isolation of myeloid cells¹⁸. To induce hypoxia *in vivo*, mice were placed in a special chamber in which N₂ and O₂ were injected to achieve an O₂ concentration of $8 \pm 0.1\%$. This was controlled by the Oxycycler hydraulic system (Model A44x0; Biospherix) and ANA-Win2 software (Version 2.4.17; Watlow Anafaze). Control mice were kept in the same room under normal atmospheric O₂ and were exposed to the same level of noise and light during each experiment. After 24 h of normoxia or hypoxia, mice were killed and their livers and brains were rapidly removed and frozen in liquid N₂ or OCT with a solid CO₂/2-methylpropan-1-ol bath.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 November 2007; accepted 7 March 2008.

Published online 23 April 2008.

- Wang, G. L. & Semenza, G. L. General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. *Proc. Natl Acad. Sci. USA* **90**, 4304–4308 (1993).
- Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
- Semenza, G. HIF-1, O₂, and the 3 PHDs: how animal cells signal hypoxia to the nucleus. *Cell* **107**, 1–3 (2001).

4. Schofield, C. J. & Ratcliffe, P. J. Oxygen sensing by HIF hydroxylase. *Nature Rev. Mol. Cell Biol.* **5**, 343–354 (2004).
5. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
6. Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF- κ B activity. *Annu. Rev. Immunol.* **18**, 621–663 (2000).
7. Häcker, H. & Karin, M. Regulation and function of IKK and IKK-related kinases. *Sci. STKE* **357**, re13 (2006).
8. Cummins, E. P. *et al.* Prolyl hydroxylase-1 negatively regulates I κ B kinase- β , giving insight into hypoxia-induced NF κ B activity. *Proc. Natl Acad. Sci. USA* **103**, 18154–18159 (2006).
9. Walmsley, S. R. *et al.* Hypoxia-induced neutrophil survival is mediated by HIF-1 α -dependent NF- κ B activity. *J. Exp. Med.* **201**, 105–115 (2005).
10. Belaiba, R. S. *et al.* Hypoxia up-regulates HIF-1 α transcription by involving PI-3 kinase and NF κ B in pulmonary artery smooth muscle cells. *Mol. Biol. Cell* **18**, 4691–4697 (2007).
11. Carbia-Nagashima, A. *et al.* RSUME, a small RWD-containing protein, enhances SUMO conjugation and stabilizes HIF-1 α during hypoxia. *Cell* **131**, 309–323 (2007).
12. Paul, S. A., Simons, J. W. & Mabeesh, N. J. HIF at the crossroads between ischemia and carcinogenesis. *J. Cell. Physiol.* **200**, 20–30 (2004).
13. Ryan, H. E., Lo, J. & Johnson, R. S. HIF-1 α is required for solid tumor formation and embryonic vascularization. *EMBO J.* **17**, 3005–3015 (1998).
14. Iyer, N. V. *et al.* Cellular and developmental control of O₂ homeostasis by hypoxia-inducible factor 1 α . *Genes Dev.* **12**, 149–162 (1998).
15. Cramer, T. *et al.* HIF-1 α is essential for myeloid cell-mediated inflammation. *Cell* **112**, 645–657 (2003).
16. Barnes, P. J. & Karin, M. Nuclear factor- κ B: a pivotal transcription factor in chronic inflammatory diseases. *N. Engl. J. Med.* **336**, 1066–1071 (1997).
17. Peyssonnaud, C. *et al.* Cutting edge: Essential role of hypoxia inducible factor-1 α in development of lipopolysaccharide-induced sepsis. *J. Immunol.* **178**, 7516–7519 (2007).
18. Greten, F. R. *et al.* NF- κ B is a negative regulator of IL-1 β secretion as revealed by genetic and pharmacological inhibition of IKK β . *Cell* **130**, 918–931 (2007).
19. Peyssonnaud, C. *et al.* HIF-1 α expression regulates the bactericidal capacity of phagocytes. *J. Clin. Invest.* **115**, 1806–1815 (2005).
20. Makino, Y. *et al.* Hypoxia-inducible factor regulates survival of antigen receptor-driven T cells. *J. Immunol.* **171**, 6534–6540 (2003).
21. Maeda, S., Kamata, H., Luo, J. L., Leffert, H. & Karin, M. IKK β couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* **121**, 977–990 (2005).
22. Schoch, H. J., Fischer, S. & Marti, H. H. Hypoxia-induced vascular endothelial growth factor expression causes vascular leakage in the brain. *Brain* **125**, 2549–2557 (2002).
23. Riboldi, E. *et al.* Cutting edge: proangiogenic properties of alternatively activated dendritic cells. *J. Immunol.* **175**, 2788–2792 (2005).
24. Störkebaum, E., Lambrechts, D. & Carmeliet, P. VEGF: once regarded as a specific angiogenic factor, now implicated in neuroprotection. *BioEssays* **26**, 943–954 (2004).
25. Chen, L. W. *et al.* The two faces of IKK and NF- κ B inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia–reperfusion. *Nature Med.* **9**, 575–581 (2003).
26. Jain, S., Maltepe, E., Lu, M. M., Simon, C. & Bradfield, C. A. Expression of ARNT, ARNT2, HIF1 α , HIF2 α and Ah receptor mRNAs in the developing mouse. *Mech. Dev.* **73**, 117–123 (1998).
27. Elson, D. A., Ryan, H. E., Snow, J. W., Johnson, R. S. & Arbeit, J. M. Coordinate up-regulation of hypoxia inducible factor (HIF)-1 α and HIF-1 target genes during multi-stage epidermal carcinogenesis and wound healing. *Cancer Res.* **60**, 6189–6195 (2000).
28. Li, Z. W. *et al.* The IKK β subunit of I κ B kinase (IKK) is essential for nuclear factor κ B activation and prevention of apoptosis. *J. Exp. Med.* **189**, 1839–1845 (1999).
29. Li, Q., Van Antwerp, D., Mercurio, F., Lee, K. F. & Verma, I. M. Severe liver degeneration in mice lacking the I κ B kinase 2 gene. *Science* **284**, 321–325 (1999).
30. Leach, R. M. & Treacher, D. F. Oxygen transport-2. Tissue hypoxia. *Br. Med. J.* **317**, 1370–1373 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.R. and M.G. were supported by a postdoctoral fellowship from the Spanish Ministry of Education and Science. Work in the laboratories of M.K., R.S.J., K.A., V.N. and G.G.H. was supported by grants from the National Institutes of Health. M.K. is an American Cancer Society Research Professor.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.K. (karinoffice@ucsd.edu).

METHODS

Quantitative RT-PCR. Total RNA was extracted with Trizol (Invitrogen) and reverse-transcribed with random hexamers and SuperScript II Kit (Invitrogen). Real-time PCR was performed with SYBR Green PCR Master Mix Kit (Applied Biosystems). The following primer pairs were used: VEGF, 5'-CCACGTC-AGAGAGCAACATCA-3' and 5'-TCATTCTCTCTATGTGCTGGCTTT-3'; PGK, 5'-GGAAGCGGGTCGTGATGA-3' and 5'-GCCTTGATCCTTTGG-TTGTTG-3'; GLUT-1, 5'-CATCCTTATTGCCAGGTGTT-3' and 5'-GAAGACGACACTGAGCAGCAGA-3'; iNOS, 5'-GGCAGCCTGTGAGACCTTG-3' and 5'-CATTGGAAGTGAAGCGTTTCG-3'; COX-2, 5'-GTGGAAA-AACCTCGTCCAGA-3' and 5'-GCTCGGCTTCCAGTATTGAG-3'; HIF-1 α , 5'-ACAAGTCACCACAGGACAG-3' and 5'-AGGGAGAAAATCAAGTCG-3'; HIF-2 α , 5'-CAACCTGCAGCCTCAGTGTATC-3' and 5'-CACCACGTCG-TTCTTCTCGAT-3'; 18S rRNA, 5'-CGCCGCTAGAGGTGAAATTCT-3' and 5'-CGAACCTCCGACTTTCGTTCT-3'.

Immunoblotting. Whole-cell extracts were obtained by lysing cells in 1% SDS, 10 mM Tris-HCl pH 7.4. Cytoplasmic and nuclear extracts were obtained as described². Proteins were separated by SDS-PAGE and detected by immunoblotting. Blots were incubated with antibodies against phosphorylated IKK- α / β , phosphorylated I κ B α , IKK- α , IKK- β , I κ B α , RelA and histone H3 (all from Santa Cruz Biotechnology), actin (Sigma), HIF-1 α , HIF-2 α and HIF-1 β (Novus).

Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) was performed with ChIP-IT Express Kit (Active Motif) in accordance with the manufacturer's instructions. Chromatin was precipitated with RelA antibodies (Santa Cruz Biotechnology). Samples were analysed by PCR. The murine HIF-1 α and actin promoters were amplified with the primer pairs 5'-CACCCCATCTCCTTTCTCT-3' and 5'-GGGTTCTCTCGAGATCCAATG-3', and 5'-TGCACTGTGCGGCGAAGC-3' and 5'-TCGAGCCATAAAAGGCAA-3', respectively.

Luciferase assay. A murine HIF-1 α -luciferase reporter, pHIF-1 α /Luc, was kindly provided by S. W. Ebbinghaus. pHIF-1 α /Luc was co-transfected with the internal control pRL-TK into either *IKK β ^{+/+}* or *IKK β ^{-/-}* MEFs with Lipofectamine 2000 (Invitrogen). Luciferase activity was measured with the Dual-luciferase reporter assay system (Promega). Results are presented as relative reporter activity after normalization to the internal control pRL-TK.

Statistical analysis. Results are expressed as means and s.e.m. A Stat View II (Abacus Concepts) statistical package was used for all analyses: multiple groups were compared by one-factor analysis of variance, followed by Fisher's protected least-squares difference to assess specific group differences.

LETTERS

Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation

Robert K. McGinty¹, Jaehoon Kim², Champak Chatterjee¹, Robert G. Roeder² & Tom W. Muir¹

Numerous post-translational modifications of histones have been described in organisms ranging from yeast to humans¹. Growing evidence for dynamic regulation of these modifications, position- and modification-specific protein interactions, and biochemical crosstalk between modifications has strengthened the 'histone code' hypothesis, in which histone modifications are integral to choreographing the expression of the genome^{1,2}. One such modification, ubiquitylation of histone H2B (uH2B) on lysine 120 (K120) in humans³, and lysine 123 in yeast⁴, has been correlated with enhanced methylation of lysine 79 (K79) of histone H3 (refs 5–8), by K79-specific methyltransferase Dot1 (KMT4)^{9–11}. However, the specific function of uH2B in this crosstalk pathway is not understood. Here we demonstrate, using chemically ubiquitylated H2B, a direct stimulation of hDot1L-mediated intranucleosomal methylation of H3 K79. Two traceless orthogonal expressed protein ligation (EPL) reactions were used to ubiquitylate H2B site-specifically. This strategy, using a photolytic ligation auxiliary and a desulphurization reaction, should be generally applicable to the chemical ubiquitylation of other proteins. Reconstitution of our uH2B into chemically defined nucleosomes, followed by biochemical analysis, revealed that uH2B directly activates methylation of H3 K79 by hDot1L. This effect is mediated through the catalytic domain of hDot1L, most likely through allosteric mechanisms. Furthermore, asymmetric incorporation of uH2B into dinucleosomes showed that the enhancement of methylation was limited to nucleosomes bearing uH2B. This work demonstrates a direct biochemical crosstalk between two modifications on separate histone proteins within a nucleosome.

It has been proposed that uH2B may induce H3 K79 methylation directly, either by altering chromatin structure and therefore nucleosomal accessibility, or through the recruitment of enzymatic function^{12,13}. However, the possibility that one or more additional factors may be required to translate the effect of uH2B into heightened Dot1 methyltransferase activity is equally likely^{14–16}. To decipher the role of uH2B in H3 K79 methylation, it is necessary to generate or purify homogeneously ubiquitylated H2B. *In vivo*, ubiquitin ligases catalyse the site-specific condensation of the carboxy-terminal carboxylic acid of ubiquitin and the ϵ -NH₂ of the target lysine in H2B, forming an isopeptide bond^{4–6,17,18}. *In vitro* reconstitution of H2B with these ubiquitin ligases and associated factors allows the production of uH2B in limited quantities⁶. Because of its natural abundance, uH2B can also be purified from endogenous sources³. However, heterogeneity due to the presence of additional modifications may complicate biochemical analyses. We decided to use EPL technology to ubiquitylate H2B regioselectively, thus bypassing the requirement for the complex cellular ubiquitylation machinery and ensuring chemical homogeneity. EPL allows the formation of an amide bond between two polypeptides of recombinant and synthetic origins, one

containing a C-terminal- α -thioester, and the other, an amino-terminal cysteine (Supplementary Fig. 1) (ref. 19).

In designing a semi-synthesis of site-specifically ubiquitylated H2B, we imagined it would be necessary to link three polypeptide building blocks together covalently, one of synthetic and two of recombinant origins (Fig. 1a). Owing to the absence of native cysteines in both H2B and ubiquitin, such a scheme requires the use of two traceless ligation strategies to generate native uH2B (Fig. 1b). In the first ligation, we reasoned that a photolytically removable, thiol-bearing ligation auxiliary could be used. We have previously demonstrated that this auxiliary allows for the site-specific ubiquitylation of a lysine side chain in model peptides²⁰. Adapting this approach to the production of a ubiquitylated protein requires the incorporation of additional functionality to facilitate a second regioselective EPL reaction. With this in mind, peptide 1 was synthesized corresponding to residues 117–125 of H2B bearing both the ligation auxiliary, attached to the ϵ -NH₂ of K120, and an A117C mutation. Orthogonal side-chain protection of K120 allowed the ligation auxiliary to be coupled to this amino group through a glycyl linker (Supplementary Fig. 2). This linker will eventually become Gly76 of ubiquitin and, importantly, allows us to exploit the Gly-Gly sequence at the C terminus of ubiquitin as an optimal junction for auxiliary-mediated ligation (Fig. 1b). Also critical to the synthetic design was transient protection of the N-terminal cysteine in 1, which precluded unwanted double ligation of ubiquitin. After examining several possibilities, we settled on the photoremovable S-(*o*-nitrobenzyl) group for this purpose, as this proved to be completely stable during the course of the first ligation reaction and was easily removed by photolysis.

In the first step of the synthesis, peptide 1 was ligated to ubiquitin(1–75)- α -thioester, 2, produced by thiolysis of a corresponding intein fusion protein (Supplementary Fig. 3), to give ubiquitylated peptide, 3 (Fig. 2a, panel i, and Supplementary Fig. 4). Subsequent ultraviolet irradiation of 3 led to efficient removal of both the ligation auxiliary and the cysteine protecting group to give deprotected branched protein 4 (Fig. 2a, panel ii, and Supplementary Fig. 4). This intermediate product was then ligated to recombinant H2B(1–116)- α -thioester, 5 (Supplementary Fig. 3), to give uH2B(A117C), 6 (Fig. 2a, panel iii and Supplementary Fig. 4). In the final step, Raney-nickel-mediated desulphurization²¹ was used to convert the single cysteine residue in branched protein 6 to the native alanine residue present in uH2B (Fig. 2a, panel iv). Under optimized conditions, this reduction was found to be highly specific for cysteine desulphurization over methionine, thereby affording native ubiquitylated H2B, 7 (Fig. 2b). The overall yield of the semi-synthesis (that is, steps i–iv in Fig. 1b) was excellent (20%). After purification, semi-synthetic uH2B was successfully incorporated into core histone octamers with wild-type recombinant H2A, H3 and H4 (Fig. 2c;

¹Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, New York 10065, USA. ²Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, New York 10065, USA.

Supplementary Fig. 5). Additional histone octamers were formed containing a recombinant H3(K79R) mutant, either alone or in combination with uH2B. These octamers were used to reconstitute chemically defined mononucleosomes with a 147 base pair (bp) region of the 601 nucleosomal targeting sequence (Fig. 2d) (ref. 22).

With modified nucleosomes in hand, we next turned our attention to exploring the effect of uH2B on Dot1 methyltransferase activity. A Flag-tagged version of human Dot1 (Flag-hDot1L) was isolated from an insect cell overexpression system and found to methylate unmodified histone octamers (Supplementary Figs 6 and 7). However, no activity was observed on H3 alone, (H3/H4)₂ tetramers or on mononucleosomes and chromatinized plasmids assembled with recombinant unmodified histones (Fig. 3a and Supplementary Fig. 7). This suggests that the presence of nucleosomal DNA represses the activity of hDot1L *in vitro*. This is in stark contrast to yeast methyltransferase, Dot1, which exhibits greater activity on recombinant nucleosomes than octamers²³. We predicted that ubiquitylation of H2B would stimulate hDot1L methyltransferase activity in the context of nucleosomes. To test this hypothesis, a ³H-SAM methyltransferase assay was performed using Flag-hDot1L and our chemically defined nucleosomes. Robust methyltransferase activity was observed on mononucleosomes containing uH2B, whereas no radioactivity was detected on unmodified mononucleosomes (Fig. 3a). This activity was specific for K79 of H3 as hDot1L was incapable of methylating nucleosomes

containing both uH2B and H3(K79R). Importantly, the level of methylation activity observed on ubiquitylated mononucleosomes was far greater than that observed on unmodified histone octamers (Supplementary Fig. 7). hDot1L also methylated a chromatinized plasmid containing uH2B (Supplementary Fig. 8). These results establish a direct biochemical connection between ubiquitylated H2B and H3 K79 methylation by hDot1L.

Although H3 K79 and H2B K120 reside in separate polypeptides, their side chains are located in close proximity to one another on the face of the nucleosome²⁴, posing a structural basis for a crosstalk pathway. Indeed, docking of the structure of the catalytic domain of hDot1L onto the mononucleosome structure, positions the catalytic domain adjacent to the site of H2B ubiquitylation²⁵. Therefore, the catalytic domain of hDot1L, containing residues 1–416, was purified from an *Escherichia coli* expression system to interrogate its role in uH2B-dependent H3 methylation (Supplementary Fig. 6). Like our observations with full-length hDot1L, a significant enhancement in activity of the catalytic domain of hDot1L was measured on mononucleosomes containing uH2B, compared with unmodified mononucleosomes (Fig. 3b). However, unlike the full-length enzyme, hDot1L(1–416) also exhibited some, albeit minimal, methyltransferase activity on unmodified mononucleosomes.

H2B ubiquitylation has been correlated with increased levels of di- and trimethylation of H3 K79 in humans^{5,6} and yeast²⁶, respectively.

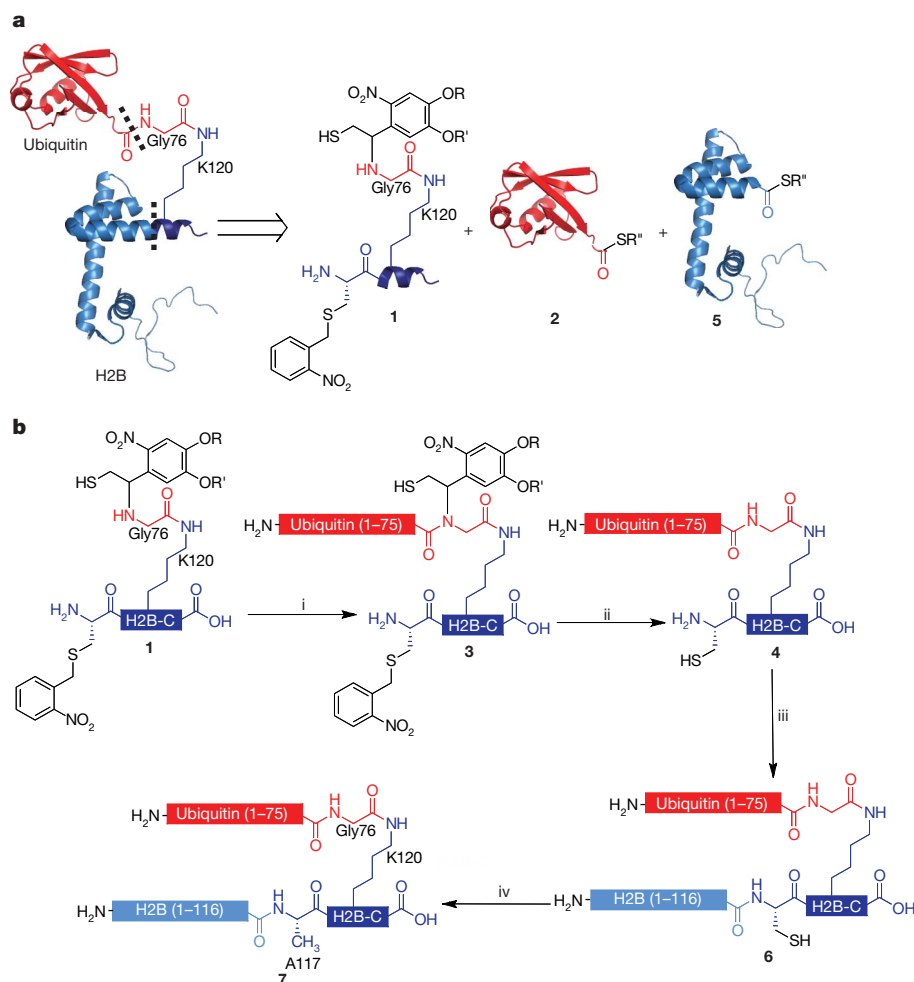


Figure 1 | Semi-synthesis of ubiquitylated H2B. **a**, Retrosynthetic analysis of uH2B synthesis. uH2B was generated by a three-piece ligation with the following polypeptides: auxiliary-linked synthetic peptide containing residues 117–125 of H2B and bearing an A117C mutation, H2B-C, **1**; recombinant ubiquitin(1–75)- α -thioester, **2**; and recombinant H2B(1–116)- α -thioester, **5**. Dashed lines indicate junctions formed by EPL reactions.

b, Synthetic scheme for the ubiquitylation of H2B. i, EPL was used to ligate peptide **1** to protein **2**, forming branched protein **3**. ii, Ligation product **3** was irradiated with 365-nm light, yielding protein **4**. iii, Ligation of protein **4** to protein **5**, forming uH2B(A117C), **6**. iv, Raney nickel desulphurization of protein **6**, forming uH2B, **7**. R = CH₂CH₂CH₂C(O)NH₂CH₃; R' = CH₃; R'' = CH₂CH₂SO₃H.

Therefore, we examined the degree of methylation occurring in our assays. In-gel trypsin digestion of H3, followed by matrix-assisted laser desorption/ionization (MALDI)-mass spectrometry showed some monomethylation and no di- and trimethylation of unmodified mononucleosomes, in assays performed with either full-length hDot1L or the catalytic domain alone (Fig. 3c, d, top panels). However, robust mono- and dimethylation of nucleosomes containing uH2B was observed in both cases (Fig. 3c, d, bottom panels). No evidence of trimethylation was observed in our assays, which is consistent with mass spectrometry analysis of H3 K79 methylation in human cell lines²⁷.

We have demonstrated that uH2B is required for effective methylation of mononucleosomes by hDot1L. However, *in vivo* levels of H3 K79 methylation vastly exceed those of H2B ubiquitylation^{3,27}. This raises the question: can uH2B enhance hDot1L-mediated methylation in an internucleosomal fashion? To address this possibility, a strategy for the creation of asymmetric dinucleosomes²⁸ was adopted. Dinucleosomes with asymmetric incorporation of uH2B and H3(K79R) were produced by ligating uniquely assembled 5' and 3' mononucleosomes with complementary DNA overhangs (Fig. 4a). The use of non-palindromic complementary overhangs resulted in only the desired heterodinucleosome products being formed (Supplementary Fig. 9). Unmodified, wild-type nucleosomes ligated to nucleosomes containing uH2B and H3(K79R) showed no increased hDot1L-mediated methylation compared with unmodified, wild-type nucleosomes ligated to nucleosomes bearing H3(K79R) alone (Fig. 4b, lanes 2 and 3). However, hDot1L was capable of methylating dinucleosomes, as evidenced by methylation of nucleosomes containing uH2B ligated to nucleosomes containing uH2B and H3(K79R) (Fig. 4b, lane 4). It could be demonstrated that

no significant histone shuffling was occurring during these assays because an unligated mixture of nucleosomes containing both uH2B and H3(K79R) and unmodified, wild-type nucleosomes showed no methylation above that observed with dinucleosomes containing only H3(K79R) (Fig. 4b, lanes 1 and 5). These results strongly suggest that efficient methylation of H3 K79 requires the presence of uH2B in the same nucleosome. Therefore, we propose that most H3 K79 methylated nucleosomes *in vivo* are likely to have at one time carried uH2B, which has since been removed by deubiquitylating enzymes or histone replacement.

The simplest explanation that accounts for the intranucleosomal stimulatory effect of uH2B on hDot1L-mediated methylation is that uH2B recruits hDot1L to nucleosomes. However, the presence of a large excess of free ubiquitin had only a modest effect on the extent of hDot1L-mediated H3 K79 methylation in uH2B containing nucleosomes (Supplementary Fig. 10). The catalytic domain of hDot1L has previously been demonstrated to bind to unmodified nucleosomes in electrophoretic mobility shift assays²⁵. We observed no significant difference in the recruitment of hDot1L(1–416) to mononucleosomes as a function of ubiquitylation, based on gel-shift analysis, and only a modest difference in gel-shift competition assays, over a broad range of concentrations (Fig. 4c and Supplementary Fig. 11). This establishes that hDot1L(1–416) is able to bind to both ubiquitylated and unmodified nucleosomes under the conditions of our

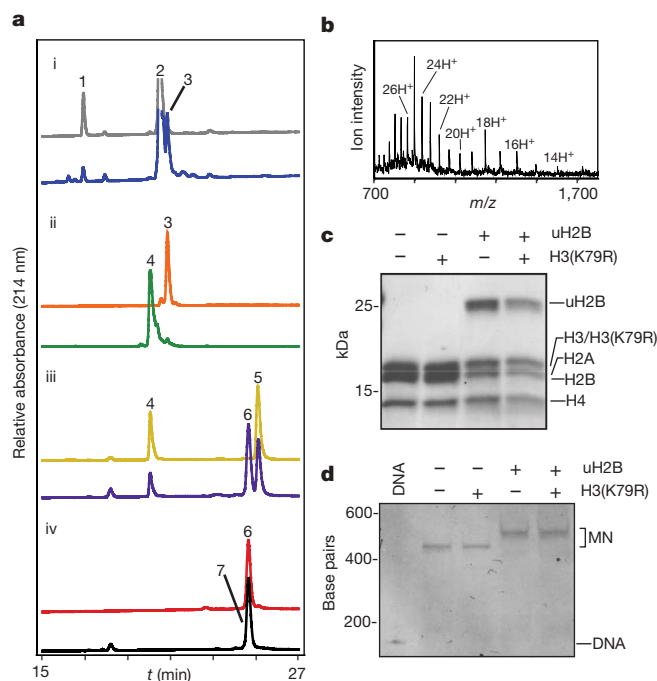


Figure 2 | Semi-synthesis of uH2B and incorporation into chemically defined histone octamers and nucleosomes. **a**, Superimposed reverse-phase high-performance liquid chromatography (RP-HPLC) chromatograms of starting materials (top trace) and products (bottom trace) of each pair, for reactions i–iv described in Fig. 1b. **b**, Electrospray ionization mass spectrometry (ESI-MS) spectrum of purified uH2B. **7**. Charge states are labelled. $(M + H)^+$ observed: $22,366 \pm 4$ Da (s.d.); expected: 22,365 Da. **c**, Reconstituted octamer samples were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and staining with Coomassie blue. **d**, Ethidium-bromide-stained native gel of reconstituted mononucleosomes (MN). Octamers and mononucleosomes contain recombinant, unmodified histones except where otherwise noted.

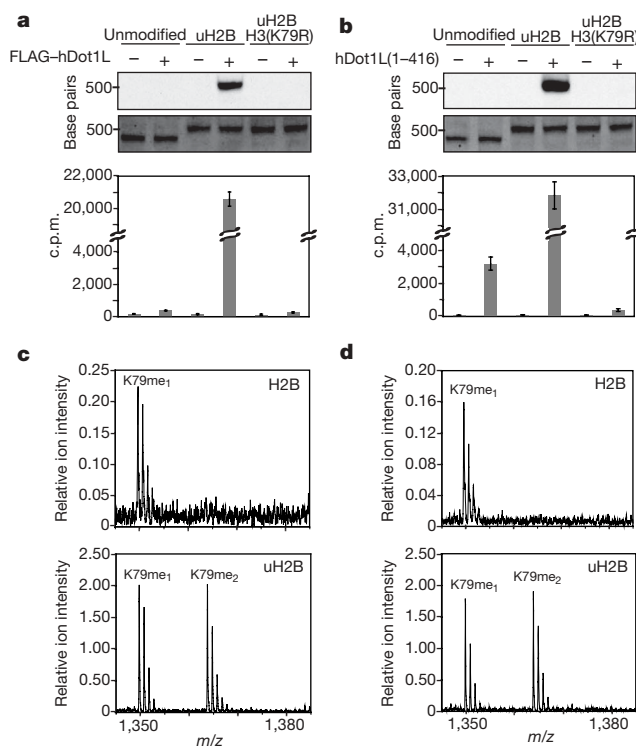


Figure 3 | Effects of uH2B on intranucleosomal methylation of H3 K79 by hDot1L. **a**, **b**, hDot1L methyltransferase assay using chemically defined nucleosomes. Assays were performed on mononucleosomes with ^3H -SAM and either full-length Flag-hDot1L, **a**, or a glutathione S-transferase (GST) fusion of the catalytic domain, hDot1L(1–416), **b**. Assay samples were separated on a native gel and stained with ethidium bromide (middle panel) before probing for ^3H -methyl incorporation by fluorography (top panel). Quantification of methylation was performed by filter binding followed by liquid scintillation counting (bottom panel). Error bars, s.e.m. ($n = 4$ –7). **c**, **d**, Characterization of degree of methylation by mass spectrometry. In-gel trypsin digestion of H3 from methyltransferase assays using Flag-hDot1L, **c**, and hDot1L(1–416), **d**, followed by MALDI-mass spectrometry. Ion intensities of methylated peptides (residues 73–83) are scaled relative to an internal standard. Assays were performed using unmodified mononucleosomes (top panel) and nucleosomes containing uH2B (bottom panel).

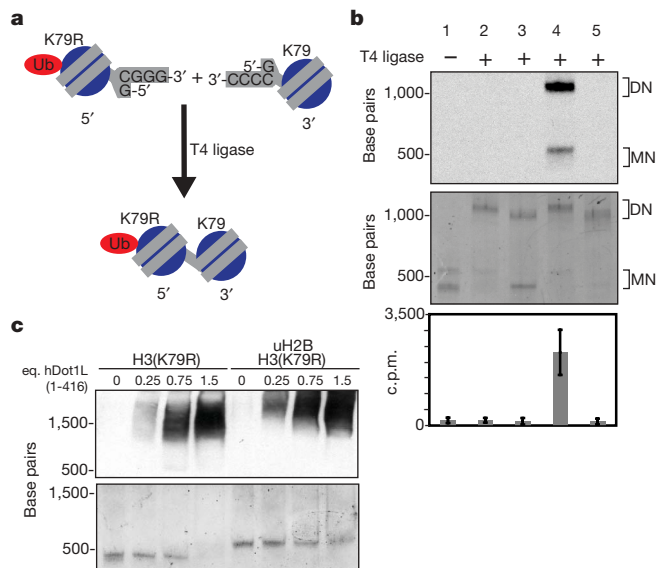


Figure 4 | Methyltransferase assays on dinucleosomes using full-length hDot1L. **a**, 5' and 3' mononucleosomes (MN) were reconstituted separately using DNA containing complementary overhangs. Assembly of 5' and 3' nucleosomes with different histones, followed by ligation, resulted in asymmetric dinucleosome (DN) formation. **b**, ^3H -SAM methyltransferase assays performed with Flag-hDot1L and dinucleosomes were separated on a native gel and stained with ethidium bromide (middle panel) before ^3H -methyl detection by fluorography (top panel). Filter-binding and liquid scintillation counting was used to quantify ^3H -methyl incorporation (bottom panel). Lanes (nucleosomes listed as [5'] – [3'] and [5'] + [3'], for ligated and unligated nucleosomes, respectively): 1 = [uH2B/H3(K79R)] + [unmodified]; 2 = [uH2B/H3(K79R)] – [unmodified]; 3 = [H3(K79R)] – [unmodified]; 4 = [uH2B/H3(K79R)] – [uH2B]; 5 = [uH2B/H3(K79R)] – [H3(K79R)]. Error bars, s.e.m. ($n = 8$). **c**, Electrophoretic-mobility shift assay performed with hDot1L(1–416) and indicated nucleosomes. Nucleosomes were incubated with increasing amounts of hDot1L(1–416), separated by native gel electrophoresis and stained with ethidium bromide (bottom panel). A western blot was performed against hDot1L using an anti-GST antibody (top panel; eq., molar equivalent). H3(K79R) was used in these assays to eliminate potential affinity differences introduced by disproportionate methylation.

assays, even though efficient methyltransferase activity is only observed on ubiquitylated nucleosomes. Thus it is likely that uH2B, possibly through an allosteric mechanism, allows hDot1L to bind nucleosomes in a catalytically competent manner. Additional investigations, perhaps involving structural analysis, will be necessary to reveal more about this phenomenon.

In summary, we have reported the chemical synthesis of a site-specifically ubiquitylated H2B. This strategy will allow the attachment of ubiquitin and Gly-Gly containing ubiquitin-like modifiers to other histones and non-histone proteins, at native and non-native sites. Chemical control around the isopeptide bond presents the option of incorporating chemical probes to study context-specific ubiquitin attachment and removal. Incorporation of our uH2B into chemically defined mono- and dinucleosomes revealed a direct stimulation of intranucleosomal methylation of H3 K79 by hDot1L. We anticipate that this combination of semi-synthetic protein chemistry to create engineered chromatin substrates, and their subsequent biochemical analysis, will play an important role in the hypothesis-driven dissection of the mechanisms underlying the 'histone code'².

METHODS SUMMARY

Protein semi-synthesis. Peptide 1 was synthesized by 9-fluorenylmethoxycarbonyl (Fmoc)-based solid-phase protocols. Orthogonal 4-methyltrityl (Mtt) protection allowed attachment of the glycyl linker and auxiliary to the side chain of K120. Recombinant proteins 2 and 5 were expressed in *E. coli* as the

corresponding intein fusions and converted into the reactive α -thioesters by thiolysis with 2-mercaptoethane sulphonic acid (MES). All protein ligation reactions were performed under aqueous conditions at pH 7.8 in the presence of an excess of MES. Photolysis of intermediate 3 was performed for 4 h with 365 nm (4 mW cm^{-2}) radiation.

Nucleosome formation. Semi-synthetic uH2B and recombinant *Xenopus* histones were reconstituted into octamers by dialysis from 6 M guanidinium chloride into 2 M NaCl and purified by size exclusion chromatography. Mononucleosomes were formed by combining purified octamers with 147 bp dsDNA corresponding to the 601 targeting sequence, followed by stepwise dilution into low salt buffer. Dinucleosomes were generated by T4 ligase-mediated ligation of pre-formed mononucleosomes bearing complementary 5' and 3' overhangs.

Methyltransferase assays. Nucleosomes (4.25 pmol) were combined with purified recombinant full-length hDot1 or its catalytic domain (0.12–0.18 pmol) and incubated with ^3H -SAM (14 pmol) for 1 h. Mixtures were then resolved by native gel electrophoresis and visualized by ethidium bromide staining and by fluorography. Quantification of K79 methylation was determined by filter binding followed by scintillation counting.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 December 2007; accepted 12 March 2008.

Published online 30 April 2008.

- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- West, M. H. P. & Bonner, W. M. Histone 2B can be modified by the attachment of ubiquitin. *Nucleic Acids Res.* **8**, 4671–4680 (1980).
- Robzyk, K., Recht, J. & Osley, M. A. Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501–504 (2000).
- Kim, J., Hake, S. B. & Roeder, R. G. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol. Cell* **20**, 759–770 (2005).
- Zhu, B. *et al.* Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol. Cell* **20**, 601–611 (2005).
- Ng, H. H., Xu, R. M., Zhang, Y. & Struhl, K. Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. *J. Biol. Chem.* **277**, 34655–34657 (2002).
- Briggs, S. D. *et al.* Trans-histone regulatory pathway in chromatin. *Nature* **418**, 498 (2002).
- Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).
- Ng, H. H. *et al.* Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* **16**, 1518–1527 (2002).
- van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).
- Henry, K. W. & Berger, S. L. Trans-tail histone modifications: wedge or bridge? *Nature Struct. Biol.* **9**, 565–566 (2002).
- Sun, Z. W. & Allis, C. D. Ubiquitylation of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**, 104–108 (2002).
- Krogan, N. J. *et al.* The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* **11**, 721–729 (2003).
- Ezhkova, E. & Tansey, W. P. Proteasomal ATPases link ubiquitylation of histone H2B to methylation of H3. *Mol. Cell* **13**, 435–442 (2004).
- Lee, J. S. *et al.* Histone crosstalk between H2B monoubiquitylation and H3 methylation mediated by COMPASS. *Cell* **131**, 1084–1096 (2007).
- Hwang, W. W. *et al.* A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol. Cell* **11**, 251–266 (2003).
- Wood, A. *et al.* Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol. Cell* **11**, 267–274 (2003).
- Muralidharan, V. & Muir, T. W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nature Methods* **3**, 429–438 (2006).
- Chatterjee, C., McGinty, R. K., Pellois, J. P. & Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem. Int. Edn Engl.* **46**, 2814–2818 (2007).
- Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526–533 (2001).
- Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
- Sawada, K. *et al.* Structure of the conserved core of the yeast Dot1p, a nucleosomal histone H3 lysine 79 methyltransferase. *J. Biol. Chem.* **279**, 43296–43306 (2004).

24. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
25. Min, J., Feng, Q., Li, Z., Zhang, Y. & Xu, R. M. Structure of the catalytic domain of human Dot1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **112**, 711–723 (2003).
26. Shahbazian, M. D., Zhang, K. & Grunstein, M. Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. *Mol. Cell* **19**, 271–277 (2005).
27. Garcia, B. A. *et al.* Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* **282**, 7641–7655 (2007).
28. Zheng, C. & Hayes, J. J. Intra- and inter-nucleosomal protein–DNA interactions of the core histone tail domains in a model system. *J. Biol. Chem.* **26**, 24217–24224 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge H. Deng and J. Fernandez at The Rockefeller University Proteomics Resource Center for mass spectrometric analysis of methylated peptides. We thank T. J. Richmond for donating the 12_177_601 plasmid. We thank C. D. Allis for contributing the *Xenopus* histone plasmids for recombinant histone expression. We thank Y. Zhang for donating a plasmid containing hDot1L. We thank B. R. Rosenberg for assistance with phosphorimaging. We thank C. D. Allis, J. Tanny, and K. P. Chiang for discussions. This work was funded by the US National Institutes of Health. R.K.M. was supported by National Institutes of Health MSTP grant GM07739. J.K. was supported by the LLS SCOR grant.

Author Contributions R.K.M., J.K. and C.C. did the experimental work; all authors performed project planning, data analysis and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.W.M. (muirt@rockefeller.edu).

METHODS

General methods. Amino-acid derivatives, pre-loaded Wang resin and coupling reagents were purchased from Novabiochem. *E. coli* BL21(DE3) and pLysS cells were purchased from Novagen. ³H-S-adenosyl methionine, goat anti-GST antibody, Amplify solution, pGex2T vector, Sephacryl S-200 resin, GSTPrep FF 16/10 column and HiTrap SP HP 1 ml column were obtained from GE Healthcare. HRP-conjugated rabbit anti-goat antibody was obtained from Dako. Restriction enzymes, polynucleotide kinase, T4 ligase, pTXB1 vector and chitin resin were obtained from New England Biolabs. Criterion 15% and 18% Tris HCl and Criterion 5% TBE gels were purchased from Biorad. Centricons were purchased from Sartorius. PCR purification and gel extraction kits were purchased from Qiagen. All other chemical reagents were purchased from Sigma-Aldrich or Fisher Scientific. Analytical and semi-preparative scale RP-HPLC were performed on a Hewlett-Packard 1100 series instrument using Vydac C18 columns (4 mm × 150 mm; 10 mm × 250 mm) at 1 and 4 ml min⁻¹, respectively. Unless otherwise noted, all analytical gradients were 0–73% B over 30 min (A: 0.1% trifluoroacetic acid (TFA) in water; B: 90% acetonitrile, 0.1% TFA in water). Preparative and process scale RP-HPLC were performed on a Waters DeltaPrep 4000 system connected to a Waters 486 tunable detector using Vydac C18 columns (22 × 250 mm; 50 × 250 mm) at 15 and 30 ml min⁻¹, respectively. Size-exclusion and ion-exchange chromatography were performed on an AKTA FPLC system from GE Healthcare equipped with a P-920 pump and a UPC-900 monitor. ESI-MS was performed on a Sciex API-100 single quadrupole mass spectrometer. Primer synthesis and DNA sequencing were performed by Integrated DNA Technologies and Genewiz, respectively.

Peptide synthesis. The sequence corresponding to residues 117–125 of *Xenopus* H2B with an A117C replacement was synthesized on pre-loaded Wang resin using manual solid-phase peptide synthesis with an Fmoc N² protection strategy and using 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) for amino-acid activation. Standard ^tbutyl side-chain protection was used throughout with the following exceptions: the ε-amino group of K120 was protected with the Mtt group, and the thiol group of C117 was protected with an *o*-nitrobenzyl group (Supplementary Fig. 2). N²-(^tbutoxycarbonyl)-S-(*o*-nitrobenzyl)-L-cysteine (Boc-Cys(ONB)) used in the peptide synthesis was prepared as previously described^{29,30}. The glycyl linker and ligation auxiliary were installed on the solid phase as follows (Supplementary Fig. 2): (i) the Mtt group on K120 was deprotected by successive incubations of the peptidyl-resin with 1% TFA in DCM containing 1% triisopropylsilane (TIS) for 10 min intervals, until no yellow colour evolved; (ii) bromoacetic acid (222 mg, 1.6 mmol) was triple coupled to the ε-NH₂ of K120 with DIC (1.6 mmol) for 3 × 1 h to yield peptidyl-resin **9**; (iii) the ligation auxiliary 4-[4-(1-amino-2-*t*-butyldisulphanyl-ethyl)-2-methoxy-5-nitro-phenoxy]-N-methylbutyramide²⁰ (32.6 mg, 76 μmol) and DIEA (1.4 mmol) were added to resin **9** (120 mg, 76 μmol) suspended in 600 μl DMF. The alkylation was allowed to proceed for 72 h at room temperature, after which the resin was dried. After cleavage from the resin with TFA:TIS:H₂O (95:2.5:2.5) for 3 h, peptide **10** was purified by RP-HPLC on a preparative scale using a 28–38% B gradient over 45 min, yielding 10.8 mg peptide. Peptide **10** was characterized by ESI-MS ((M + H)⁺ observed: 1,606.8 Da; expected: 1,607.9 Da) (Supplementary Fig. 2).

Preparation of ubiquitin(1–75) and H2B(1–116)-α-thioesters. Ubiquitin (1–75)-MES (**2**) was prepared as previously described²⁰. A truncated *Xenopus* H2B gene, containing residues 1–116, was amplified by PCR using primers H2B-FP (5'-GGAATTCATATGCTAGCCAGCCCAAGTCCGCTCCAGCC-CCG-3') and H2B116-RP (5'-GGTGGTTGCTCTTCCGCACTTGGTGCCCTCGGACAC-3') and a *Xenopus* H2B expression plasmid³¹ as a template. After digestion by *Nde*I and *Sap*I, the fragment was ligated into a similarly digested pTXB1 vector, and the resulting plasmid, pRMH2B-N, which encodes H2B (1–116) fused at its C terminus to the GyrA intein and a chitin-binding domain, was verified by DNA sequencing.

E. coli BL21(DE3) cells transformed with pRMH2B-N were grown in Luria-Bertani medium at 37 °C until mid-log phase, and protein expression was induced by the addition of 0.5 mM IPTG and allowed to continue at 25 °C for 16 h. After harvesting the cells by centrifugation at 6.8 kg for 15 min, the cell pellet was resuspended in lysis buffer (50 mM Tris, 200 mM NaCl, 1 mM EDTA, pH 7.5) and frozen at –80 °C. Thawed cells were lysed by passage through a French press and the insoluble material was removed by centrifugation at 26,000g for 30 min. The supernatant was filtered and incubated overnight at 4 °C with chitin resin (35 ml) pre-equilibrated in lysis buffer. The resin was washed with 200 ml of wash buffer 1 (50 mM Tris, 200 mM NaCl, 1 mM EDTA, pH 7.2) and 700 ml wash buffer 2 (50 mM Tris, 200 mM NaCl, 1 mM EDTA, pH 7.4). The resin was then incubated with cleavage buffer (50 mM Tris, 200 mM NaCl, 1 mM EDTA, 100 mM MESNa, pH 7.4) for 70 h, resulting in thiolysis of the intein fusion, forming H2B(1–116)-MES, **5**. The column was

eluted and the resin was washed with 2 × 25 ml cleavage buffer. The thiolysis reaction was repeated and the combined elution fractions further purified by preparative RP-HPLC using a 42–52% B gradient over 45 min, yielding 4–5 mg of lyophilized protein per litre of culture. The identity of the purified protein was verified as **5** by ESI-MS ((M + H)⁺ observed: 12,991 ± 3 Da; expected: 12,991 Da) (Supplementary Fig. 3). Mass spectrometry indicated that the non-native N-terminal methionine used for expression of **5** was processed during recombinant expression, leaving the native N-terminal sequence.

Expressed protein ligation reaction 1. The ligation reaction between peptide **1** and ubiquitin(1–75)-MES, **2**, was performed using conditions similar to those previously optimized for model reactions²⁰. In a typical reaction, purified peptide **10** (1.1 mg, 0.69 μmol) was dissolved in 180 μl of buffer containing 3 M guanidinium chloride, 300 mM sodium phosphate pH 7.8, 50 mM tris(2-carboxyethyl)phosphine (TCEP), and incubated at room temperature for 30 min to remove the S-^tbutyl protection group on the ligation auxiliary. The resulting reduced peptide **1** was added to ubiquitin(1–75)-MES, **2** (17.1 mg, 1.98 μmol) and dissolved in ligation buffer (3 M guanidinium chloride, 300 mM sodium phosphate, 100 mM MES, pH 7.8). The reaction volume was increased to 950 μl with ligation buffer and the pH was adjusted to 7.8 using NaOH. After 120 h at 4 °C, the reaction was quenched with 1 ml of 50% HPLC buffer B containing 100 mM TCEP. Ligation product **3** was purified using preparative HPLC with a 32–42% B gradient over 45 min, yielding 4.0 mg of lyophilized protein. ESI-MS was used to verify the identity of the ligation product ((M + H)⁺ observed: 10,008 ± 1 Da; expected: 10,009 Da) (Supplementary Fig. 4).

Photolysis. In a typical reaction, ligation product **3** (3.5 mg, 0.35 mmol) was dissolved in 1.5 ml of photolysis buffer (25% HPLC buffer B containing 10 mM semicarbazide, 10 mM DTT, 10 mM ascorbic acid and 2 mM cysteine)³⁰. The resulting solution was irradiated at 365 nm for 4 h using a collimated light source from Oriel equipped with a 200 W Hg lamp. Irradiance (4 mW cm⁻²) was measured by using a model 840-c monochromatic photometer. Selective irradiation at 365 nm was achieved by using an analytical line filter (9.4 nm bandwidth) obtained from Oriel. Irradiation effected removal of the ligation auxiliary and the *o*-nitrobenzyl protecting group, forming branched protein **4**. Semi-preparative RP-HPLC purification of protein **4** was accomplished using a 0–73% B gradient over 45 min, yielding 2.5 mg of lyophilized protein **4**. Removal of the two photolytic groups was verified by ESI-MS ((M + H)⁺ observed: 9,547 ± 2 Da; expected: 9,548 Da) (Supplementary Fig. 4).

Expressed protein ligation reaction 2. In a typical reaction, photolysis product, **4** (2.5 mg, 0.25 μmol), and H2B(1–116)-MES, **5** (5.0 mg, 0.38 μmol), were dissolved in ligation buffer to a final volume of 225 μl. The pH of the resulting solution was increased to 7.8 with NaOH and the reaction was allowed to proceed for 78 h before quenching with 225 μl of 50% HPLC buffer B containing 100 mM TCEP. The ligation product, **6**, was purified using semi-preparative RP-HPLC with a 42–52% B gradient over 45 min, yielding 3.0 mg of uH2B(A117C), **6**. The identity of the ligation product was verified by ESI-MS ((M + H)⁺ observed: 22,395 ± 3 Da; expected: 22,397 Da) (Supplementary Fig. 4).

Desulphurization. Raney nickel reduction was used to convert C117 of protein **6** to the native alanine^{21,32}. In a typical reaction, uH2B(A117C), **6**, (1.3 mg, 58 nmol) was dissolved in 3 ml desulphurization buffer (6 M guanidinium chloride, 200 mM sodium phosphate, 35 mM TCEP). Raney nickel was prepared by adding 200 mg NaBH₄ to a stirred solution of 1.2 g of nickel acetate in 6 ml of water. After 5 min, the Raney nickel was filtered, washed with 200 ml water and added to the solution of uH2B(A117C), **6**. The reaction progress was followed by HPLC and ESI-MS. An identical amount of fresh Raney nickel was added after 6 h and the reaction was found to be complete at 8.5 h. (Note: much longer reaction times (over 24 h) led to a second desulphurization reaction on methionine.) The Raney nickel was pelleted by centrifugation and washed with 4 × 0.5 ml desulphurization buffer. The reaction supernatant and washes were combined, added to an equivalent volume of 50% HPLC buffer B, and purified using semi-preparative RP-HPLC with a 42–52% B gradient over 45 min, yielding 1.1 mg of uH2B, **7**. uH2B was characterized by ESI-MS ((M + H)⁺ observed: 22,366 ± 4 Da; expected: 22,365 Da).

Recombinant histone preparation. Recombinantly expressed *Xenopus* histones H2A, H2B, H3 and H4 were prepared similarly to previously described³³. The three N-terminal residues of H2B (PEP) were added to the *Xenopus* H2B expression plasmid³¹. The DNA encoding residues 1–125 of H2B was amplified by PCR using primers H2B-FP described above and H2B-RP (5'-CGGGATCCTTACTGGCGCTGGTGTACTTG-3') and the H2B plasmid described above as a template. After digestion with *Nde*I and *Bam*HI, the H2B gene was ligated into a similarly digested pET vector from the *Xenopus* H2A expression plasmid³¹. An H3(K79R) point mutation was generated with a QuikChange II XL kit (Stratagene) using primers H3K79R-FP (5'-GCTCAGGACTTCAGGACCGA-CCTGCGC-3') and H3K79R-RP (5'-GCGCAGGTCGGTCTGAAGTCCT-

GAGC-3') and the *Xenopus* H3 expression plasmid³¹ as a template. The resulting histone expression plasmids were verified by DNA sequencing.

For protein expression, *E. coli* BL21(DE3)pLysS cells, transformed with the appropriate histone expression plasmid, were grown in 612 × TY media at 37 °C until mid-log phase, and protein expression was induced by the addition of 0.5 mM IPTG at 37 °C for 2–3 h. Cells were harvested and lysed as described above for proteins 2 and 5. The insoluble fractions of the bacterial lysates were washed twice with 20 ml wash buffer (20 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5) and once with 20 ml of Triton wash buffer (20 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 1% Triton, pH 7.5). DMSO (1 mL) was then added to the pellets and after 15 min, 15–50 ml of extraction buffer (7 M guanidinium chloride, 20 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5) was added and the suspension stirred for a further 15–30 min before clearing by centrifugation at 26,000g for 30 min. The supernatants were purified using a Sephacryl S-200 column (approximately 1 l bed volume), eluting with extraction buffer. Purified fractions were combined, dialysed into water containing 2 mM DTT and lyophilized in aliquots suitable for octamer formation. H3(K79R) was further purified using process RP-HPLC with a 45–65% B gradient over 60 min. The identity of all purified histones was verified by ESI-MS (Supplementary Fig. 5).

Preparation of DNA for nucleosome and dinucleosome formation. A plasmid containing 12 copies of 177 bp of the 601 sequence³⁴ was purified using a Qiagen Plasmid Giga kit. The 177-bp repeat was digested from the vector using *EcoRV* sites flanking the segment. The desired segment was selectively precipitated by incrementally increasing the concentration of PEG-6000 from 4% to 8.5%, followed by centrifugation at 26,000g to separate precipitated DNA. *SacI* digestion and gel purification using a QIAquick Gel Extraction Kit afforded a 177-bp fragment of the 601 sequence, 1_177_601. PCR amplification of the central 147-bp region of 1_177_601 was accomplished using primers 147-FP (5'-CTGGAG-AATCCCGGTGCCGAGG-3') and 147-RP (5'-ACAGGATGTATATATCTG-ACACG-3') and gel-purified 1_177_601 as a template. Purification of the PCR product using a QIAquick PCR purification kit yielded 1_147_601.

5' and 3' DNA fragments for dinucleosome formation were generated using PCR amplification as described above. The 5' and 3' fragments were designed to contain 3' and 5' *DraIII* restriction sites, respectively, allowing for the formation of complementary overhangs. The DNA segment for 5' mononucleosome formation, 601-5', was generated using the primers 147-FP and 5'-RP (5'-ATTGAGCACCCCGTGGGATCTTACATGCACAGGATG-3') and 1_177_601 DNA as the template. The DNA segment for 3' mononucleosome formation, 601-3', was generated using 3'-FP (5'-ATTGAGCACGGGTGCGGCCGC-CCTGGAG-3') and 147-RP primers. The PCR-amplified fragments (110 µg) were digested with *DraIII* (700 U) using NEB Buffer 3 (1 ml) for 6 h at 37 °C, followed by purification using a QIAquick PCR purification kit. When ligated, the resulting 601-5' and 601-3' sequences connect two identical 147-bp regions of the 601 sequence, separated by a 30-bp linker.

Nucleosome formation. Histone octamers were formed as previously described³³. Briefly, individual histones were dissolved in unfolding buffer (7 M guanidinium chloride, 20 mM Tris, 10 mM DTT, pH 7.5) at approximately 4 mg ml⁻¹. Histones were combined in equimolar amounts (combined protein ranged from 0.75 to 12 mg), and the solution was diluted to 1 mg ml⁻¹ with unfolding buffer. The resulting mixture was dialysed into refolding buffer (2 M NaCl, 10 mM Tris, 1 mM EDTA, 1 mM DTT, pH 7.5) (three changes of 2 l each). Crude octamer assemblies were concentrated in Vivaspin 2 and 20 centricons (3–10 kDa molecular-weight cutoff (MWCO)) and purified using a Superdex 200 10/300 column, eluted with refolding buffer. Octamer quality was verified by 18% SDS-PAGE, followed by staining with Coomassie blue. Octamer samples were stored at –20 °C in 50% glycerol.

Mononucleosomes were formed using a previously described stepwise dilution procedure³⁵. Briefly, octamers and DNA (1_147_601, 601-5' or 601-3') were combined in 10 µl high salt refolding buffer to a final concentration of 3 µM. After incubation at 37 °C for 15 min, 3.3 µl of dilution buffer 1 (10 mM HEPES, 1 mM EDTA, 0.5 mM PMSF, pH 7.9) was added and the temperature was dropped to 30 °C. Further dilutions of 6.7, 5, 3.6, 4.7, 6.7, 10, 30 and 20 µl, respectively, were then performed every 15 min. A final dilution with 100 µl of dilution buffer 2 (10 mM Tris, 1 mM EDTA, 0.1% NP-40, 5 mM DTT, 0.5 mM PMSF, 20% glycerol, pH 7.5) was performed. After an additional 15 min, the nucleosomes were concentrated using Vivaspin 500 centricons (3–10 kDa MWCO) at 4 °C. Nucleosome formation was verified by separation on a Criterion 5% TBE gel run in 0.5× TBE, followed by staining with ethidium bromide. The chromatinized plasmid was assembled using a previously described procedure³⁶ and plasmid³⁷. Micrococcal nuclease assays were used to verify equivalent chromatin reconstitution with uH2B and H2B (Supplementary Fig. 8).

Dinucleosome formation. Dinucleosome ligations were performed as previously described²⁸. In a typical ligation reaction, 5' and 3' mononucleosomes (6.25 pmol) were combined in 200 µl of 1× ligation buffer (New England Biolabs) and 3,600 U of T4 ligase were added. After 1 h at room temperature, the ligated nucleosomes were concentrated using Vivaspin 500 centricons (3–10 kDa MWCO) and dinucleosome formation was verified using a 5% TBE gel as described above.

Preparation of Flag-hDot1L and hDot1L(1–416). The cDNA for hDot1L was a gift from Y. Zhang. After modification to include an N-terminal Flag-epitope, the gene was subcloned into the pFASTBAC1 vector (Invitrogen) and the baculovirus was generated according to the manufacturer's instructions. Sf9 cells were infected with the baculovirus and the resulting cell extracts were subjected to standard purification procedures as described below. Infected Sf9 cells were collected, resuspended in lysis buffer (20 mM Tris, 500 mM NaCl, 4 mM MgCl₂, 0.4 mM EDTA, 20% glycerol, 1 mM PMSF, 2 mM DTT, pH 7.9), and disrupted with a dounce homogenizer. After removal of cell debris by centrifugation, the supernatant was adjusted to 0.1% NP-40 and 300 mM NaCl by dilution with 20 mM Tris containing 10% glycerol. The resulting solution was incubated with M2-agarose beads (Sigma). After washing extensively (20 mM Tris-HCl, 150 mM NaCl, 2 mM MgCl₂, 0.2 mM EDTA, 15% glycerol, 0.1% NP-40, 1 mM PMSF, 1 mM DTT, pH 7.9), the bound Flag-hDot1L was eluted with Flag peptide (Supplementary Fig. 6).

The catalytic domain of hDot1L encoding residues 1–416 was amplified from MSCN-hDot1L (a gift from Y. Zhang) using primers 416-FP (5'-CGGGATCC-CATCACCATCATCATCATGCGGGAAGAGCTGGAGCTG-3') and 416-RP (5'-GGAATTCCTACTTCTTGGGCGCGCCGCGC-3'). The resulting sequence was digested with *Bam*HI and *Eco*RI and ligated into a similarly digested pGexTEV vector (pGex2T with the thrombin recognition sequence replaced with a TEV sequence). The resulting vector, pGexTEV-hDot1L416, was verified by DNA sequencing. *E. coli* BL21(DE3) cells transformed pGexTEV-hDot1L416, were grown in 6 l LB media at 37 °C until mid-log phase, and protein expression of the hDot1L(1–416) GST fusion protein was induced by the addition of 0.5 mM IPTG at 18 °C for 18 h (ref. 25). Cells were harvested and lysed as described above. Cleared cell lysates were continuously loaded onto a GST-Prep FF 16/10 column at 1 ml min⁻¹ for 3 h. The column was washed with 200 ml of 20 mM Tris, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, pH 7.5 and eluted with 200 ml of the same buffer containing 5 mM glutathione. Fractions deemed pure by SDS-PAGE were stored in 20% glycerol at –80 °C. Impure fractions were diluted with an equal volume of start buffer (20 mM Tris, 1 mM EDTA, 2 mM DTT, pH 7.5) and loaded onto a HiTrap SP HP 1-ml column. After washing with 10 ml start buffer, hDot1L(1–416) was eluted with a 10-ml linear gradient from start buffer to end buffer (20 mM Tris, 1 mM EDTA, 2 mM DTT, 1 M NaCl, pH 7.5). Pure fractions were stored at –80 °C in 20% glycerol (Supplementary Fig. 6).

Methyltransferase assays. Mono- and dinucleosomes (4.25 pmol) were combined with Flag-hDot1L (0.12 pmol) or hDot1L(1–416) (0.18 pmol) in 25 µl of assay buffer (20 mM Tris, 4 mM EDTA, 0.5 mM DTT, 1 mM PMSF, pH 7.9). To initiate the reaction, ³H-S-adenosyl methionine (SAM) (1 µCi, 14.25 pmol) was added and the reaction was allowed to proceed at 30 °C for 1 h. After 1 h, 20 µl of each sample were separated on a 5% TBE gel as described above and stained with ethidium bromide. After visualization, the gel was fixed with 25% isopropanol, 10% acetic acid in water, followed by a 30 min incubation with Amplify solution. Dried gels were visualized by fluorography. The remaining 5 µl of reaction mixture were spotted on Whatman p81 filter paper, washed 3 × 10 min with sodium carbonate, pH 9, and dried. Filter papers were added to 5 ml Ready Safe Liquid Scintillation Cocktail (Beckman Coulter). The samples were vortexed for 10 s and counted using a LKB Wallac 1209 RackBeta Primo Liquid scintillation counter.

Mass spectrometric analysis of methyltransferase assays. Methyltransferase assays were performed similar to that described above but on twice the scale using cold SAM (1 mM). Assays were incubated at 30 °C for 3 h before concentration in a Vivaspin 500 centricon (10 kDa MWCO), separated on a Criterion 15% Tris HCl gel, and stained with Coomassie blue. The Coomassie-stained H3 bands were excised from the gel and destained with 200 mM ammonium bicarbonate. Tryptic digestion was initiated with the addition of 25 ng µl⁻¹ Sequence Grade Modified Trypsin (Promega) in ammonium bicarbonate buffer. The protein was digested for at least 16 h at 37 °C. The digestion products were mixed with 0.5 µl of 10 mg ml⁻¹ α-cyano-4-hydroxysuccinamic acid in 50% acetonitrile containing 0.1% TFA, and applied to a MALDI plate. MALDI mass spectra were recorded with a PerSeptive Voyager-DE STR MALDI time-of-flight mass spectrometer operated in the reflectron mode. Ion intensity of the tryptic fragment of H3 containing residues 40–49 was used to scale the intensities of the methylated peptides for comparison of separate samples.

Electrophoretic mobility-shift assay. Mononucleosomes (3.0 pmol), hDot1L (1–416) (0–4.5 pmol) and cold SAM (10 nmol) were combined in 30 μ l of assay buffer. After 1 h at 30 °C, samples were separated on a 5% TBE gel as described above and visualized with ethidium bromide staining. Proteins were transferred to a PVDF membrane and a western blot was performed against GST (goat anti-GST used at 1:1,000).

For radiolabelled gel-shift competition experiments, 1_147_601 was end-labelled with 32 P using 32P-ATP (Perkin Elmer) and polynucleotide kinase. Radiolabelled nucleosomes were prepared as described above at 2 μ M starting concentrations. Radiolabelled nucleosomes (0.64 pmol), hDot1L(1–416) (3.2 pmol) and cold nucleosomes (0–3.2 pmol) were incubated and separated as described above. Dried gels were imaged by phosphorimaging on a Typhoon 8400 (GE Healthcare).

29. Wang, S. S. *et al.* Facile synthesis of amino acid and peptide esters under mild conditions via cesium salts. *J. Org. Chem.* **42**, 1286–1290 (1977).
30. Smith, A. B., Savinov, S. N., Manjappara, U. V. & Chaiken, I. M. Peptide-small molecule hybrids via orthogonal deprotection-chemoselective conjugation to cysteine-anchored scaffolds. *Org. Lett.* **4**, 4041–4044 (2002).
31. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–311 (1997).
32. Pentelute, B. L. & Kent, S. B. Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. *Org. Lett.* **9**, 687–690 (2007).
33. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119**, 1–16 (1999).
34. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* **327**, 85–96 (2003).
35. Owen-Hughes, T. *et al.* Analysis of nucleosome disruption by ATP-driven chromatin remodelling complexes. *Methods Mol. Biol.* **119**, 319–331 (1999).
36. Ito, T. *et al.* ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* **13**, 1529–1539 (1999).
37. An, W., Kim, J. & Roeder, R. G. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell* **117**, 735–748 (2004).

naturejobs

**JOBS OF
THE WEEK**

I remember a 1980s political-science lecture in which US President Ronald Reagan's concept of a 'flexible freeze' to rein in the US budget deficit was mocked. The professor rightly labelled the phrase as "weasel words" and an example of orwellian doublespeak — it's a paradox both to freeze a budget and allow it room to grow flexibly.

So it was with some trepidation that I read the European Commission's new policy on creating more stable science jobs in the European Union (EU) using a concept it calls "flexicurity". The goals of this curious word sound noble: to create greater job security, but allow scientists to change jobs more easily. The policy advocates job security by eliminating short-term contracts and flexibility by making grants more portable.

Unlike Reagan's flexible freeze, these two goals aren't mutually exclusive. But under Europe's multi-tiered funding system, it might be difficult to achieve on a large scale. The EU can mandate its policies only on funds that it distributes, which make up only a fraction of research costs for most member states. It cannot tell individual state funders, such as the Max Planck Society in Germany or the UK Medical Research Council, what strings to attach to their grants.

Like the term flexible freeze, flexicurity hides the real challenge. In both cases, it's the same thing: a realistic budget to match the goal. The EU wants member states to spend 3% of their gross domestic product on research and development (R&D) by 2010. With less than two years to go, the average R&D expenditure lingers at around 2.5%. Without increased funding, there can be neither broad flexibility nor job security for European scientists. The current economic problems in the United States, which resemble those of the Reagan era (high budget and trade deficits), have kept research budgets relatively flat over the past several years. This might offer a boost to the EU's goal of attracting back some of the 100,000 or so EU-born scientists now working in the United States. Ironically, round two of the flexible freeze in the United States may help Europe accomplish the aims of flexicurity.

Paul Smaglik moderates the *Naturejobs* Nature Network career-advice forum.

CONTACTS

Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:

Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718

Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com



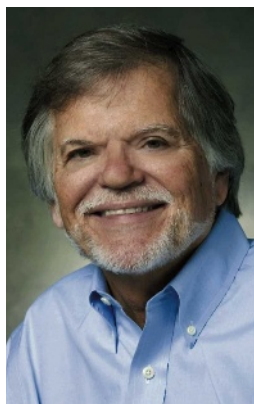
BEYOND ROCKET SCIENCE

Huntsville, Alabama, the original home of NASA and of military weapons development, makes a move into biology. **Paul Smaglik** reports from Rocket City.

In the midst of 14,000 hectares, surrounded by chain-link fences topped with razor wire and shielded from sight by a veil of trees, sits a 100-metre-tall corrugated metal tower, scarred with rust and clad in scaffolding. This is the Saturn V Dynamic Test Stand at NASA's Marshall Space Flight Center in Huntsville, Alabama. The tower tested the rockets that powered the Apollo mission to the Moon and those used for the US space-shuttle programme; its last tests were for the shuttle *Enterprise* in 1978–79. It has sat unused since 1979 and was designated a National Historic Landmark in 1985.

The tower — one of the tallest buildings in this mid-sized southern city — will soon see service again. It will help NASA scientists stress-test the next generation of rockets the United States plans to send to the Moon. Bringing this site back into service is akin to awakening a sleeping giant — perhaps an apt metaphor for Huntsville itself.

The city will see thousands of jobs open up over the next five years, and only some are related to space science and military technology. Last month, the HudsonAlpha Institute for Biotechnology formally opened its doors to reveal a facility that holds about 500 researchers. The institute, founded by Jim Hudson, who sold his Huntsville reagents company Research Genetics to Invitrogen of Carlsbad, California, in 2000, will combine academic research, education and incubation for applied research companies. It has plans for two more buildings that would house another 1,000 people. The institute is split between academic research and biotech start-ups. The academic arm has five principal investigators, with plans to add three or four more next year and another two or three the year after. As the academic side grows, so too, it is hoped, will the biotech.



Jim Hudson (top), founder of biotechnology institute HudsonAlpha, and John Horack of NASA both see a more diverse future for research in Huntsville.

NASA's switch in emphasis from shuttles to Moon missions using the ARES rocket will bring a couple of thousand more jobs to Huntsville. And the US Army's closure of some bases elsewhere in the country will benefit Huntsville by bringing in up to 5,000 jobs to the army base there. This is in addition to several hundred new jobs with aerospace firm Boeing, long established in Huntsville, and wireless-networking firm AdTran. Employees of these companies, as well as recent returnees to the area, envisage greater interaction between the private R&D sector, NASA and the army, creating more companies and providing more jobs.

Huntsville was not always a hub for aerospace research. In 1949, it was a sleepy cotton town of 12,500 people. After the Second World War, the US government secretly brought Wernher von Braun, the architect of Germany's V-2 rocket programme, to the United States. At Huntsville, von Braun developed the American intercontinental ballistic missile, was the first director of NASA's Marshall Space Flight Center and designed the Saturn V launch vehicle, which took the Apollo spacecraft to the Moon. The city grew exponentially and attracted scientists and engineers from around the world.

Now Huntsville is poised for another boost, motivated by the memory of rocket science but propelled by biology, says Ethan Hadley, vice-president of Huntsville's Chamber of Commerce. "Jim Hudson's inspiration came out of the space programme," Hadley says. Hudson, an enthusiastic but self-effacing entrepreneur, admits it is "intimidating" competing with major biotech hubs such as San Diego, Boston and the San Francisco Bay area. "I will get into trouble if I said biotech here was non-existent," Hudson says. But he touts Huntsville's perks: a high concentration

R. CUMMINS/CORBIS

HUDSONALPHA INST. BIOTECHNOL.

NASA/MSFC

of engineers, mathematicians and computer scientists who can easily cross over to biological applications.

Hudson started out in the foundry business, working with his father. But intrigued by molecular biology in the 1980s, he sold the business, enrolled in a master's programme and started a reagents business while finishing his thesis. After he graduated, Hudson moved the company from his university lab bench to the back of a Huntsville beauty supply store. As the Human Genome Project grew, he started providing biochemical reagents to labs participating in the project. The researchers would 'test drive' these reagents, tell him which ones were useful and help him tweak others. Then he marketed them around the world.

As the Human Genome Project advanced, so did Research Genetics. The company grew to 230 employees with annual revenues of US\$23 million a year before Hudson sold it. Many of the people he hired, collaborated with or networked with are now involved at HudsonAlpha or at one of the 20 or so start-ups the building holds.

Rick Myers, director of HudsonAlpha, is one such returnee. He grew up in Tuscaloosa, Alabama, and served on Research Genetics' board. At Stanford University, he helped sequence about 11% of the human genome as well as working on the genomes of about 40 other organisms. Myers — who sports stubble, a pink shirt, an earring and nary a trace of his native accent — jokes that his Alabama upbringing prepared him to return to this small military city in the Deep South. "At least I knew what I was getting into," he says.

He sees advantages in having access to Huntsville's pool of analytical talent. "We're going to need a lot of engineers and a lot of computer scientists," he says. "Those quantitative types can help solve problems and analyse data better than traditional biologists." Besides bringing nine scientists from Stanford along with him, he's hired one computer scientist from NASA.

The HudsonAlpha building's sleek pedestrian bridges encourage biotech workers to bump into academic researchers. And Myers and Hudson have networks beyond the building.

University links

The University of Alabama in Huntsville is both tapping into, and feeding, the local biotech industry. The university started an interdisciplinary PhD programme in 2001 that combines chemistry, life sciences and chemical engineering and works with local biotech companies and with NASA and Oak Ridge National Laboratory in Tennessee. "This is a unique programme in the southeast, because it's the only programme that trains PhDs in biotech," says the programme's director Joseph Ng.



Rick Myers has returned from Stanford to his native Alabama to head the HudsonAlpha institute.

In addition to course work, graduate students work 20 hours a week at companies based at the HudsonAlpha Institute and receive a stipend and tuition waivers for their efforts. The university and leaders of local biotechs have also created a non-profit entity called Partnership for Biotechnology Research. Through this, students are encouraged to develop and commercialize their own products.

The University of Alabama at Birmingham (UAB) also has a research relationship with Huntsville, strengthened through HudsonAlpha. The university is offering joint appointments to HudsonAlpha investigators. Birmingham graduate students can do their thesis research with investigators at the Huntsville institute. And the HudsonAlpha investigators will offer courses to Birmingham students.

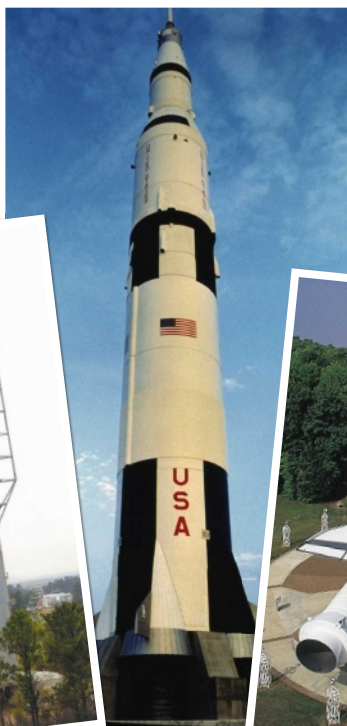
UAB's history in clinical research will also be a boon to HudsonAlpha. The university receives about \$450 million a year in grants and contracts, of which about \$200 million is from the US National Institutes of Health (NIH), making it the 20th-ranked institution in the country in NIH funding, and the recipient of the most NIH money in the southeast, says Richard Marchase, UAB's vice-president of research. One of the biggest benefits that UAB offers Huntsville in the clinical field is the half a million visits a year to its hospital by an ethnically diverse patient population.

Closer to home, structural-biology expertise at the University of Alabama in Huntsville will help HudsonAlpha avoid unnecessary costs in pursuing three-dimensional protein structure, says Edward Meehan, director of the university's structural-biology laboratory. "We'll do the NMR and X-ray crystallography," says Meehan. "So there's no crossover, no duplication." But structural-biology students at Huntsville can work with researchers both at the HudsonAlpha institute and at firms in the incubator.

Hudson also bridged the gap between academia and industry when he hired principal investigator Jian Han. Hudson met Han when the China-born scientist first started marketing diagnostic tests based on the multiplex polymerase chain reaction (PCR), which researchers use to test a patient's tissue sample against 10 or more pathogens simultaneously to determine the cause of infection and then to confirm the identity and quantity of pathogens in a sample. Hudson

approached Han to market an earlier version of one of Han's diagnostic tests for genetic markers in a Chinese population. "Jim Hudson knocked on my door one day and said 'I want to sell that product,'" Han says. "That didn't take off, but Jim said 'Why don't you come to Huntsville?'" At the time, Hudson provided free lab space and reagents. Han sold his company, Genaco, to life-sciences equipment company Qiagen in 2006 for \$22 million plus shares.

Han and Hudson's success as biotech entrepreneurs



Huntsville heritage: Saturn test stand (left), Saturn rocket (centre) and the space shuttle Pathfinder.

might also rub off on the other tenants of the HudsonAlpha building. Troy Moore, who worked for Research Genetics for 10 years, initially stayed with Invitrogen after the sale. Then he and a few others started up Open Biosystems in Huntsville in 2002. "We're not big company types," Moore says. "That's one of the great things about Huntsville. We're very entrepreneurial." Open Biosystems has grown to 68 employees and will expand as the collection of 12 million genes and hairpin RNAi constructs in its 'freezer farm' grows.

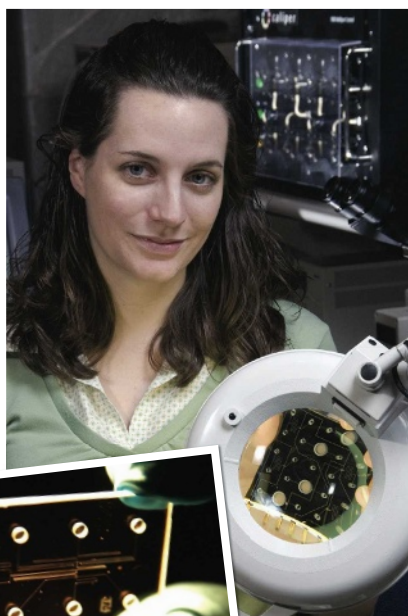
The HudsonAlpha Institute should help that growth, Moore says. Another boon is that the institute is in the Cummings Research Park, which, at around 1,500 hectares and 285 companies, is the country's second-largest science park. Most of the park's tenants have historically supported the aerospace and defence industries. More recent ones have increased the city's presence in IT and telecommunications. But investment and interest in biotech are growing.

Sowing the seed

HudsonAlpha provided initial funding for Expression Genetics, a Huntsville-based gene-therapy company using non-viral vectors encased in biodegradable polymers. Its first drug, a gene construct expressing interleukin-12 for the treatment of ovarian cancer, is in a phase I clinical trial at the University of Pennsylvania, with UAB providing analytical services. Expression Genetics expects to double in size to 30 employees in a year if the trial proceeds as expected, then to add another 15 people as the drug advances further.

Applied Genomics, with its head office in Huntsville, has taken gene- and protein-expression databases developed by Pat Brown and David Botstein at Stanford University, added antibodies to the mix and is developing diagnostics using this information. One of the company's kits helps identify which breast-cancer patients are likely to benefit from the chemotherapeutic agent tamoxifen and who might be harmed.

The company's approach marries biology, computer science and engineering. Robert Seitz, Applied Genomics' president and chief executive, thinks more



The lab-on-a-chip portable test system for detecting and identifying microbes on space-station surfaces.

such interdisciplinary mixes will come out of Huntsville, especially under Hudson's leadership. "Jim has introduced biology to a town that is by and large engineering — and is helping to bring the two together — which is where the future is," says Seitz.

Interdisciplinary NASA

That future has already arrived at NASA. The agency is developing a chip that combines microfluidics and microarray capabilities. It could help search for microorganisms in space and also has practical applications on Earth, such as detecting toxins in the air or searching hydrothermal vents or Arctic ice chunks for signs of life, says Lisa Monaco, project director of NASA's Lab-on-a-Chip Application Development programme. Scientists from many disciplines, universities and companies are involved in the project.

NASA scientists are exploring other interdisciplinary niches as well. Tom Sever, for example, is the agency's sole archaeologist. On images taken from space he has found undiscovered Mayan ruins — as well as possible evidence of a large-scale water-management system. And Richard Hoover, NASA astrobiologist, will lead an international team to Antarctica in search of new low-temperature-dwelling microorganisms that might give scientists clues to what life might be like on other planets.

Other projects, aside from the ARES rocket, are gearing up at the Marshall Space Flight Center. NASA needs thousands of scientists and engineers to prepare the James Webb Space Telescope for its 2014 launch. This replacement for the Hubble Space Telescope has a segmented mirror and detects lower-energy rays, allowing it to see deeper into space with greater resolution and clarity. Scientists at Marshall are also developing a roving science lab "roughly the size of a Mini Cooper" to survey the martian surface, says John Horack, manager of science and missions systems at the centre. As the next mission to Mars gets nearer, Horack anticipates a greater focus on planetary geology and the search for water. "In the future, I think you'll see more emphasis on Earth sciences," he says.

Steve Smith, scientific director of the US Army Aviation and Missile Command at Redstone, sees more interactions between disciplines emerging at Redstone. "Over the decades, we've looked at applying biological systems to weapons, like looking at bumble-bees and vibration sensing," says Smith, who has been with Redstone for 51 years, going back to the days when von Braun ran the labs.

Smith believes potential recruits should not be put off by the thought of moving to a southern town with a military background. NASA pioneer von Braun, ever the visionary, created an orchestra and a ballet company, and insisted that the state build a research university — now the University of Alabama in Huntsville. But the German father of American space science might not recognize the place now. With Huntsville's surprising and impressive degree of ethnic diversity, a visitor is as likely to stumble upon a German *Biergarten* as that southern staple, the Waffle House. Once more, the city seems poised to shoot for the Moon — both literally and figuratively.

Paul Smaglik is a Milwaukee-based science writer.



The Marshall centre links Earth with the International Space Station.

The invisible hand

A granted wish.

Allan M. Lees

It's funny how frequently the public's idea of something is far from its reality. Hollywood thrives on this misperception, of course, but so does science. Canvas one hundred people at random about their notions of science and you'll get a litany of descriptions involving white coats, rationality and the pursuit of knowledge. In theory science is about exploring the unknown and pressing back the boundaries of knowledge, whereas in practice it's all about writing up ever more grant applications and pressing back the boundaries of plagiarism.

The dirty little secret of science is that most of it is mediocre at best. Science is, by and large, something done by those of us who don't want to expose ourselves to the hustle and bustle of commerce. Of course, saying this out loud is heresy, and if I were still a working scientist I'd be expelled from the lab or institute in which I'd managed to create some kind of refuge from the cold hard world. But I am no longer a working scientist; at least, not directly. Now I am more of a ... well, one might say, prime mover.

Seven years ago my life was that of a typical young male scientist: most of my spare time was spent trying to impress the latest female intern with dates in cheap restaurants and much specious waffle about how one day I'd have a place for her in my own lab. But mostly it was all about writing grant applications; rewriting grant applications; waiting to hear back from the study groups that score grant applications; and then inevitably writing yet more grant applications after the first lot was rejected. I quickly learned that study groups were comprised of older scientists whose best work had been done years ago. They would reliably approve grants for barely incremental mini-steps that were essentially near-copies of what had already been done before. No really new or radical grant proposal was ever funded.

Everyone knows the story of Craig Venter, the man who first sequenced the human genome: he wrote a grant application for funds to sequence part of his own genome and had his grant rejected by the highest and most eminent scientific authorities on the grounds that such a thing was totally impossible ... and he then carried out the first sequencing just three weeks later. The big innovations such as the silicon chip, the telecommunications revolution, software,

jet transport ... everything important came out of industry, not academia, because commerce must respond to basic human needs. Research science was stagnant, conservative and dead-ended.

Until me.

One evening I was sitting in my dirty bedroom, perched on a pile of old men's magazines, typing up yet another grant application when I decided to do some basic research of my own: find out the composition of the study group that would review my latest grant proposal, discover their biases, and skew my proposal to pander to their prejudices. It's something the more senior lab members had been doing for years, of course, but no one talked about it openly. As I was reading an online article by one study-group member, I came across the phrase that changed my life, and by extension the entire future of science. The eminent professor in question was bemoaning the sheer volume of grant submissions that had to be reviewed. She said: "We have to read hundreds of grant proposals each year in order to approve a mere handful. It absorbs far too much of our time."

That banal utterance changed everything. That night I abandoned my semi-finished proposal to study the effects of α -lipoic acid on a cloned passive-aggressive subspecies of *Caenorhabditis elegans* and began to write a software program to automate the evaluation of grant proposals. It took me three months of hard work, but the result was worth it. I sent it out into the world anonymously via e-mail lists and free download sites. Naturally no one would consider using it, no one would admit to using it ... but within a few months it was evident that grant proposals were being processed far more speedily than before.

It would have been criminal negligence if I hadn't taken advantage of the opportunity. I made my first fortune with my revolutionary grant-proposal-generation service, the yin to my first program's yang. I accepted online bids and the winners received system-generated proposals that would get a 100% score when evaluated by (my) grant-evaluation software. As word spread, bids grew in number and size and

within six months I was seriously wealthy. And I could have stopped there. Bill Gates would have stopped there. I think even Sergey Brin and Larry Page would have stopped there. But I didn't. If years of reading lads' mags had taught me anything, it was that more is better.

From time to time I had updated the algorithms in my anonymous grant-evaluation program to stop other people from writing grant-proposal-generation software that would score as highly as my own and thus undermine my highly profitable monopoly.



JACEY

Now I set out to modify my program with a higher purpose in mind: by altering the scoring algorithms I could essentially determine what types of research would get funded. I, alone, could steer the direction of fundamental science across the entire scientific world. For the first time, research science could be focused on mankind's most fundamental desire. And thus I focused it.

And that is how the entire scientific establishment, some two million researchers around the world, physicists, biologists, chemists, astronomers, botanists and even neo-classical anthropologists, have come to unite as one around the study of life's single most important problem: how to achieve reliable male organ enhancement.

Allan M. Lees has been creating stories for his children since they were very little and he will continue to do so until they are old enough to steal a car and escape. Allan's very modest literary success to date includes several published stories, a now-deservedly out-of-print novel, a radio play, and many megabytes of wasted hard-drive space.